AMERICAN
STATISTICAL ASSOCIATION

NEW SERIES, No. 123. SEPTEMBER, 1918

OUR MEMORIAL VOLUME.

By Wesley C. Mitchell, *President of the American Statistical Association.*

The American Statistical Association passed its seventy-fifth birthday on the 13th of February, 1914. To celebrate that anniversary the Association decided to prepare a history of official statistics in the principal countries of the world. Mr. John Koren, then our president, undertook the arduous duty of collecting and editing the contributions.

The Memorial Volume thus planned has but recently appeared. In February, 1914, no one foresaw the tragic events which have rendered the editor's task so difficult, and have delayed its completion so long. But neither did anyone then foresee the peculiar significance which these events would lend the volume.

Official statistics, several of the contributors remind us, have had a longer and more chequered history than we commonly remember. Professor Faure, in his admirable paper upon French experience, begins with the *polyptiques* of Pepin the Short. Sir Athelstane Baines begins the English article with Domesday Book. In the centuries since these medieval beginnings there are many landmarks pointed out in the progress of European statistics. Even the writers on the newest countries have to divide their stories into "periods." But though recognized in the development of every country's statistics these "periods" seldom coincide in any two cases. Now, however, we are passing through years that will mark a turning-point in the history of statistics in every great nation in the world.

Imperious necessity is driving all nations, neutrals as well as belligerents, to organize their economic life with a new thoroughness. How much food do the people require, how much do they raise, how much must they import? Similar problems

thrust forward in respect to every great staple, in respect to railways and ships, ports and factories, capital and labor. Foreign resources are scanned not less anxiously than domestic. In every capital city there are men raging over the deficiencies of the statistical data provided by other countries. In every capital responsible statesmen are wondering how far they dare trust their own familiar figures, and demanding in haste data that have never been gathered. New statistical bureaus with inquisitorial powers and large resources are being set up to meet emergency calls. Very different from what they were in 1914 will be the statistical organizations on the return of peace.

Our Memorial Volume brings the story of statistics in most of the leading nations down to the outbreak of war. Better than any other book it sums up the accomplishments of the day that has passed. That gives it unique historical value.

But the volume has a larger service to render than that of recording what has been. It will also help in shaping what is to be in the future. Statistics is an international enterprise. Washington is concerned both with the statistical output and with the statistical methods of London, Paris, and Berlin, of Buenos Ayres, Wellington, and Tokyo. The officials of every country that is striving to improve its statistical service will wish to survey the procedure of other lands. There may be few occasions for exact imitation of a foreign model, but there are many occasions for seeking hints. Perhaps no one entrusted with a task of statistical re-organization can stop with the information found in our Memorial Volume; but no one can find a better book with which to begin.

In these days of stress the American Statistical Association has an unprecedented number of its members enrolled in the nation's service, many of them at heavy personal sacrifice. Perhaps this work for the country will involve a temporary slackening in the Association's corporate activities. But statistical methods are demonstrating so forcibly their utility as a solvent of administrative perplexities that we may feel every confidence in the future not only of the science but also of the Association that fosters it. In the meantime we are fortunate in having the Memorial Volume to remind us what the Association has accomplished in the past and what it should strive toward in the future.

THE RELATION BETWEEN BIRTH RATE AND DEATH RATE IN A NORMAL POPULATION AND THE RATIONAL BASIS OF AN EMPIRICAL FORMULA FOR THE MEAN LENGTH OF LIFE* GIVEN BY WILLIAM FARR.

By Alfred J. Lotka.

Both birth rate and death rate are, for obvious reasons, functions of the age-distribution in a given population.

A certain interest attaches to this circumstance from several points of view.

It is of interest to the statistician who seeks to "correct" the so-called "crude" death rates so as to bring them to a definite and common standard for purposes of comparison.†

Secondly, the subject is of interest in connection with certain problems of biochemistry, the growth of certain colonies of living organisms having been found to take place in accordance with the law of monomolecular reactions. As has been pointed out by Eijkman,‡ the question there arises, how this law of growth can be harmonized with the influence of the age-distribution on birth rates and death rates.§

Thirdly, the relation between age-distribution and birth and death rates is of interest in connection with the mathematical theory of evolution. For a quantitative measure of the "fitness" of a given type of organism, or of its "adaptation to given conditions," is given in its rate of growth under those conditions.‖ But this rate of growth depends, not only on the fundamental characteristics of the species, but also on a factor

* Expectation of Life at birth.

† Cf. G. H. Knibbs, The Establishment of a Series of Norms for Death Rates. Trans. XV. Internat. Congr. Hyg. and Demogr. (1912), v. VI, p. 343; Knibbs & Wickens, The Establishment and Use of Population Norms, etc. *Loc. cit.* p. 352.

A. J. Lotka. A Natural Population Norm. Jl. Washington Ac. Sc. 1913, v. 3, pp. 241, 289.

‡ Cf. C. Eijkman, On the Reaction of Velocity of Organisms. Proc. Amsterdam Ac. Sci., 1912, p. 629; Reichenbach. Zeitschr, f. Hyg. u. Inf. Krankh, 1911, v. 69, p. 171; Wo. Ostwald, Uber die Zeitlichen Eigenschaften der Entwickelungsvorgänge, Leipsic, 1908; T. B. Robertson, On the Normal Rate of Growth of an Individual, Arch. f. Entwickelungsmechanik 1908, p. 108.

§ In ordinary chemical reactions and radioactive transformations this question does not arise, as the "force of mortality" of a species of molecules is independent of the age of the molecules. See A. J. Lotka, Am. Jl. Sci., 1907, v. 24, p. 205.

‖ Cf. A. J. Lotka, Am. Jl. Science, 1907, v. 24, p. 216; Ann. Nat. Phil., 1911, p. 59; Jl. Wash. Ac. Sci., 1912, pp. 2, 49, 66; 1915, pp. 360, 397.

which, from this point of view, must be regarded as purely adventitious, namely the age-distribution among the individuals of the species.

The problem here arises of eliminating this adventitious factor from the case under consideration.

Fortunately it very largely eliminates itself by a spontaneous process: It is evident that the age distribution in a population is not fortuitous. It is capable of variation and fluctuation, but not without all rule and beyond all bounds. In point of fact, in an isolated population, where emigration and immigration are absent or negligible, the age-distribution tends to approach a certain definite normal type,* and when this is attained a relation between birth rate and death rate exists which no longer explicitly involves the age-distribution. Such a population behaves as if its birth rate and death rate were independent of the age-distribution, and grows and diminishes according to the exponential law or the law of geometric progression,† the same law which governs the progress of a monomolecular chemical reaction. There would therefore seem to be no occasion for surprise at the observation discussed by Eijkman, *loc. cit.*, that the growth of certain organisms follows this law irrespectively of the age-distribution in the colony.

While in practice the "normal" age-distribution may be more or less closely approached, it will, for obvious reasons,

* Cf. Poynting & Thomson, Text-book of Physics, Volume on Heat, 1904, p. 135; A. J. Lotka, Am. Jl. Sci., 1907, v. 24, p. 201; Science, 1907, v. 26, p. 21; Jl. Wash. Ac. Sci., 1913, pp. 241, 289; Sharpe & Lotka, Phil. Mag., April, 1911, p. 435.

† Cf. also M. Block, Traité de Statistique, 1886, p. 209; A. J. Lotka, Am. Jl. Sc., 1907, v. 24, p. 201.

It is evident that the birth rate and death rate can be *actually* independent of the age-distribution only if the force of mortality is independent of the age, *i. e.*, in the notation of equation (1) in the text, p. 123, if

$$\underset{k=\text{const.}}{p(a)} = s^{-ka}$$

When the age-distribution is fixed, and the population *behaves as if* its death rate were independent of the age-distribution, we can speak of an *equivalent constant force of mortality*, the value of which follows from equation (4), Am. Jl. Sc., 1907, v. 24, p. 200, viz:

$$\begin{aligned} d &= -\int_0^\infty c(a)\ \frac{d \log p(a)}{da}\ da \\ &= k \int_0^\infty c(a)\ da \\ &= k \end{aligned}$$

so that the equivalent constant force of mortality is numerically equal to the death rate per head.

never be exactly attained in any real population. But for the purposes of theoretical discussion, where we wish to eliminate from consideration the influence of the adventitious factor of age-distribution, it is convenient to base one's reflections on the example of an ideal population in which the "normal" age-distribution is established. This is all the more legitimate, since it has been shown that in practice this age-distribution may at times be very closely approached.*

A roughly approximate expression for the relation between the birth rate and death rate in a population which is increasing in geometrical progression (and in which, therefore, the age-distribution is normal) has been given by Bristowe.† This expression is derived on the assumption that each individual attains the mean length of life, and is quite inexact.

The exact relation between the birth rate and death rate in a normal population has been given by the author‡ in the form

$$\frac{1}{b} = \int_0^\infty e^{-ra} \quad p(a)da \tag{1}$$

where

b = birth rate per head per annum
r = natural rate of increase per head per annum
$= b - d$
d = death rate per head per annum
$p(a)$ = probability at birth of reaching age a (tabulated in the life tables and there commonly denoted by the symbol l_x).

Equation (1) expresses the relation between b and d in implicit form and is inconvenient for purposes of numerical computation.

In a later publication§ another expression was therefore derived from (1) by expanding e^{-ra} under the integral sign and integrating term by term.

The formulae thus obtained give the relation between b and r and d and r in explicit form and offer certain advantages for purposes of computation, but are still somewhat compli-

* A. J. Lotka, Jl. Wash. Ac. Sc., 1913, pp. 243, 244.

† St. Thomas' Hosp. Report, 1876, quoted in Newsholme, Vital Statistics, 1899, p. 295.

‡ Am. Jl. Sci., *loc. cit.*

§ Jl. Wash. Ac. Sci., 1913, pp. 241, 289. Equations of the form (1) have also been dealt with by Pincherle, Mem. Bologna Ac. Sc., 1887, v. 8, p. 10; Bateman, Proc. Cambridge Phil. Soc., 1910, v. 15, p. 424.

cated, since the series converge, for common values of r, at such rate as to necessitate the retention of five or six terms, and the somewhat laborious determination of a corresponding number of constants.

A very much simpler form of the relation between b and d in a normal population, which holds with a very fair degree of approximation, has now been found, suggested by some reflections on an empirical (approximate) formula for the mean length of life quoted by Newsholme and attributed by him to Farr.* At the same time, following up this clue, an insight was gained into the *raison d'être* of Farr's empirical formula.

Farr showed that during the decade 1881–1890, the birth rate and death rate in England and Wales were related to the mean length of life† L very nearly according to the equation

$$\frac{1}{3}\cdot\frac{1}{b}+\frac{2}{3}\cdot\frac{1}{d}=L \tag{2}$$

This equation is of the form

$$\frac{P}{b}+\frac{Q}{d}=L \tag{3}$$

or

$$\frac{p}{b}+\frac{q}{d}=1 \tag{4}$$

By a simple transformation, we have, from (4)

$$(b-p)\,(d-q)=pq \tag{5}$$

which would indicate that for a given value of L the relation between b and d is, as a first approximation, hyperbolic in form.

Indeed, a glance at curve I in the accompanying figure, which is drawn in accordance with the data published by the writer in his paper on *A Natural Population Norm*, confirms this indication at least qualitatively. This curve shows the graph of the relation

* Newsholme, Vital Statistics, 1889, p. 301.

† For further discussion of the relation between the mean length of life and the rate of increase of a population see W. J. Spillman, Science, 1912; republished, with editorial note by the present writer, in Sci. Am. Suppl., 1912, p. 378.

$$\frac{1}{b}=\int_0^\infty e^{-(b-d)a}\ p(a)da \tag{1a}$$

$$=\int_0^\infty e^{-ra}\ p(a)da \tag{1}$$

for the decade 1871–1880 in England and Wales.

To obtain further light on this point, we proceed as follows: Equation (1a) is an implicit relation between b and d. We may write it in the form:

$$\frac{1}{b}=F_1(r) \tag{6}$$

$$\frac{1}{d}=F_2(r) \tag{7}$$

whence, by eliminating r between (6) and (7) we should obtain

$$\frac{1}{b}=F_3\left(\frac{1}{d}\right) \tag{8}$$

Now F_1 and F_2 are known in the form of series, can, in fact, be obtained by expanding e^{-ra} under the integral sign in (1) and integrating term by term. The value of the coefficients of successive powers of r can be determined by mechanical quadrature when $p(a)$ is given in form of a life table and r is known.

We thus obtain, for England and Wales, 1881–1890

$$\frac{1}{b}=L+Ur+Vr^2+Wr^3+\ \ .\ .\ . \tag{9}$$

$$\frac{1}{d}=L+ur+vr^2+wr^3+\ \ .\ .\ . \tag{10}$$

Neglecting terms in r^2 and higher powers of r, and eliminating r between (9) and (10) we have:

$$\frac{u}{b}-\frac{U}{d}=L(u-U) \tag{11}$$

$$\frac{u}{u-U}\cdot\frac{1}{b}-\frac{U}{u-U}\cdot\frac{1}{d}=L \tag{12}$$

an equation which corresponds in form to Farr's formula (2) or its generalizations (3) and (4). This correspondence extends not only to the general linear form but also to the fact that the sum of the coefficients of $\frac{1}{b}$ and $\frac{1}{d}$ in (12) is unity, just as in Farr's formula (2).

NUMERICAL DATA.

1. Farr applies his formula to the period 1881–1890 (England and Wales).

The writer has determined by the method indicated in his paper cited above the coefficients u, U. There is first obtained

$$b = .02200 + .7130r + 6.766r^2 - 17.47r^3 + \quad . \quad . \quad . \qquad (13)$$

$$d = .02200 - .2870r + 6.766r^2 - 17.47r^3 + \quad . \quad . \quad . \qquad (14)$$

This gives, neglecting the terms in r^2 and r^3,

$$\frac{1}{d} = 45.45 + 593.0r + \quad . \quad . \quad . \qquad (15)$$

$$\frac{1}{b} = 45.45 - 1473r + \quad . \quad . \quad . \qquad (16)$$

$$.713\,\frac{1}{d} + .287\,\frac{1}{b} = 45.45 \qquad (17)$$

so that we have for the mean length of life

$$L = .713\,\frac{1}{d} + .287\,\frac{1}{b} \qquad (18)$$

The observed mean values of b and d during this period were

$$\left.\begin{aligned} b &= .03234 \\ d &= .0190 \end{aligned}\right\} \qquad (19)$$

Substituting these in (18) gives

$$L = 37.5 + 8.8 = 46.3$$

Observed	$L = 45.5$
Discrepancy	.8

Farr's formula gives

$$L = 45.6$$

an exceedingly close fit, which, however, is undoubtedly accidental.

2. For the period 1871–1880 the author's previous work furnishes detailed data for the male population.

It appeared of interest to recompute the graph of the relation between b and d by the hyperbolic formula (12) and compare it with the graph of the more exact relation (1) previously plotted.

The numerical data in this case are as follows:

$$b = .02418 + .7673r + 6.823r^2 + \quad . \quad . \quad . \tag{20}$$

$$d = .02418 - .2327r + 6.823r^2 + \quad . \quad . \quad . \tag{21}$$

This gives

$$.767\,\frac{1}{d} + .233\,\frac{1}{b} = 41.35 \tag{22}$$

This, written in hyperbolic form, gives

$$(b - .00563)\,(d - .01855) = .00010443 \tag{23}$$

or, putting

$$B = b - .00563 \tag{24}$$

$$D = d - .01855 \tag{25}$$

we have

$$BD = .00010443 \tag{26}$$

Table I below shows, for successive arbitrarily chosen values of b, the corresponding values of d calculated by (23), and, for comparison, also the corresponding values of d calculated by (1). For comparison also there are shown in the last two columns the corresponding products BD. The hyperbolic formula, of course, makes this product constant. The series formula gives a certain departure from constancy, which on inspection was found to be approximately equal to $-2r^3$.

This suggested a simple empirical formula which gives a very close approximation to the theoretically exact relation (1), namely

$$(b - .00563)\,(d - .01855) = .00010443 + 2r^3 \tag{27}$$

or

$$(b - .00563)\,(b - r - .01855) = .00010443 + 2r^3 \tag{28}$$

a relation which enables us to compute a series of values of b corresponding to given values of r. The corresponding values of d then follow directly in view of the relation $(b - d) = r$.

TABLE I.

I b	II d exact	III d hyperb.	IV* $BD \times 10^5$ exact	V* $BD \times 10^5$ hyperb.
.040	.02197	.02158	11754	10443
38	2205	2178	11329	10443
36	2217	2199	10994	10443
34	2234	2223	10752	10443
32	2256	2251	10574	10443
30	2284	2284	10454	10443
28	2322	2322	10447	10443
26	2368	2368	10450	10443
24	2423	2424	10434	10443
20	2574	2582	10332	10443
16	2796	2862	9758	10443
12	3125	3495	8090	10443
.008	3603	.06262	4143	10443

* $\begin{cases} B=b-.00563 \\ D=d-.01855 \end{cases}$

TABLE II.

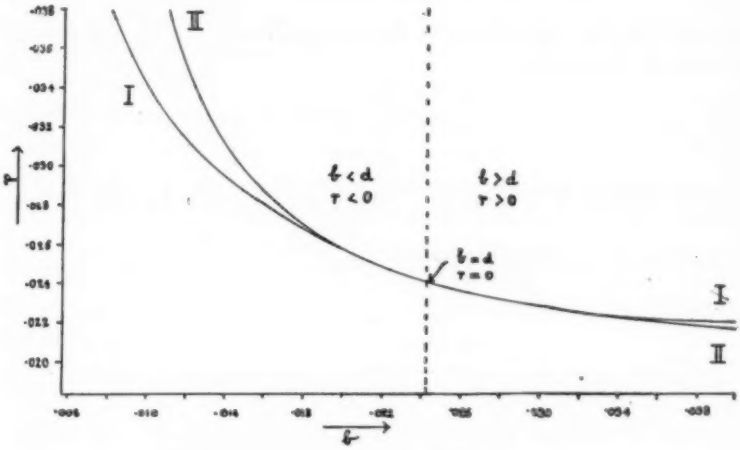
I r	II b exact	III b corr. hyp.	IV d exact	V d corr. hyp.	VI $BD \times 10^5$ exact	VII $BD \times 10^5$ corr. hyp. $=10443+2r^3 \times 10^5$
.01803	.040	.04000	.02197	.02197	11754	11615
1595	38	3798	2205	2203	11329	11258
1383	36	3599	2217	2216	10994	10972
1166	34	3400	2234	2234	10752	10760
0944	32	3201	2256	2257	10574	10609
0716	30	3002	2284	2286	10454	10516
0478	28	2801	2322	2323	10447	10465
0232	26	2600	2368	2368	10450	10446
−.00023	24	2400	2423	2423	10434	10443
− 574	20	2003	2574	2577	10332	10405
− 1196	16	1617	2796	2813	9758	10101
− 1975	12	1225	3175	3200	8408	8902
−.02803	.008	.00799	3603	3602	4143	6039

Table II above shows, for the values of r corresponding according to equation (1) to successive entries in Table I, the values of b and d calculated both by the exact formula (1) and also by the corrected hyperbolic formula (28). The last two columns show the product BD as given by the exact formula (1) and also by the corrected hyperbolic formula (28). A comparison of columns II, III of Table I shows that, except for extremely small values of b, the hyperbolic formula gives very fair agreement with the exact formula.

On the other hand columns II, III and IV, V of Table II show an almost perfect agreement, which is somewhat remarkable in view of the empirical nature of the very simple correcting term $2r^3$.

The same features are brought out graphically in the accompanying figure which needs no further explanation beyond that furnished in the caption.

The agreement between columns II, III and IV, V of Table II is so close that there is no object in showing the corresponding curves. They would practically coincide within the errors of drawing.



CURVE I. GRAPH OF $b = \dfrac{1}{\int_0^\infty e^{-(b-d)a} p(a)da}$

CURVE II. GRAPH OF $.767\ \frac{1}{b} + .233\ \frac{1}{d} = 41.35 = L$

or $(b - .00563)\ (d - .01855) = .00010443$

Note that within the region of commonly occurring values of b and d the two curves are nearly coincident.

To the left of the dotted vertical line the death rate exceeds the birth rate, so that the population is diminishing. The marked divergence of curves I and II on the extreme left is of no practical significance, since such rapid rates of decrease as here represented can not long continue, if they occur at all, in any real population.

Summary:

The implicit relation between birth rate per head b and death rate per head d

$$\frac{1}{b} = \int_0^\infty e^{-(b-d)a} \quad p(a)da \qquad \text{(I)}$$

which was developed in a previous paper, is now shown to be equivalent to the explicit hyperbolic relation

$$(b-p)\ (d-q) = pq = \text{const} \qquad \text{(II)}$$

in the neighborhood of $b-d=0$ and for values of b and d such as occur in practice.

The hyperbolic relation (II), expressed as a relation between $\frac{1}{b}$ and $\frac{1}{d}$, assumes the linear form

$$\frac{u}{u-U}\cdot\frac{1}{b}-\frac{U}{u-U}\cdot\frac{1}{d}=L \qquad \text{(III)}$$

which is of the same form as Farr's empirical rule for the mean length of life L, viz.:

$$\frac{1}{3}\cdot\frac{1}{b}+\frac{2}{3}\cdot\frac{1}{d}=L \qquad \text{(IV)}$$

The empirical constants $\frac{1}{3}$ and $\frac{2}{3}$ which occur in Farr's rule thus receive a rational interpretation.

OCCUPATIONAL PROPINQUITY AS A FACTOR IN MARRIAGE SELECTION.

By Donald M. Marvin.

The presence of a large and increasing number of women in industry raises the question of the possible influence of industry upon marriage selection. Modern social conventions are based upon the presumption that woman is shut away in the home and that man must follow her there if he wishes to see her. In entering industry in such large numbers, women face two new conditions, one negative, the other positive. They leave the home temporarily empty and they create a new social phenomenon of occupational propinquity. This, modified by the various influences of class stratification, financial status, and other forms of group cohesion involved in the present organization of society, tends to differentiate certain groups of men and women for marriage.

Industrial propinquity extends to the home and to the economic status. Even class and race lines enforce occupational cohesion. Such stratification and drawing together of certain parts of social groups must react variously upon those involved. It seems possible that friends who marry within their own occupation are not so much guided by similar tastes and backgrounds as they are driven by a new force of industrial propinquity, a force that has developed with the appearance of woman in industry.

Marriage, a matter of individual choice, if choice exists, obeys the sweeping silent force of propinquity. Women in each occupation are surrounded by the men of the same occupation. Of course they marry these men. This inevitable sequence causes no astonishment.

Today the most attractive as well as the strongest and most vigorous women are in industry. Their presence has been accepted and the taboo has been removed. The result is that men are now marrying the women whom they meet in their work. The tremendous proportions of this movement are of startling and far-reaching significance.

Today we are confronted with tremendous numbers of women in industry and this environment must react upon all phases of their lives. Not the least important aspect of this general situation is the fact that the occupation seems to select certain groups of people for acquaintance and for eventual marriage through propinquity within the shop or factory.

To the extent that women are a factor in modern industry, industry may affect their lives. Industrial propinquity must be a factor in marriage selection in proportion to the number of women employed in industry. The report of the United States Census on Occupations for 1910 shows that there are four times as many women in industry as there were in 1880. With the tremendous acceleration of this movement due to the war, it is quite possible that the number will have doubled between 1910 and 1920. There are over 200,000 women employed in the industries of Philadelphia alone. Twenty-three and four tenths per cent. of the women of the country are engaged in gainful occupations and are therefore exposed to whatever influence occupation may have upon marriage. Of the unmarried women of marriageable age in an industrial city like Philadelphia, a far larger proportion are gainfully employed. In fact, 50 per cent. of the marrying women of Philadelphia are employed at the time of marriage.

The constant mingling of men and women within an occupation causes such constant sex stimulation as to exert a determining influence upon marriage selection. The effect of a particular occupation upon a woman and upon her chance of marriage is more directly the problem of this thesis. The professional woman may serve as an example. When a woman physician marries a man physician, as seven out of ten such Bryn Mawr alumnae actually did, a new type of home environment is the result. The ethics and ideals of the profession become table talk and professional ambitions and desires are not hedged about by office hours. Such an atmosphere inevitably affects the children. If this were a matter of isolated cases, it might be dismissed as of no importance, but already it concerns half the women being married today.

The problem, then, is the extent and constancy with which industry affects marriage. The first measure of this influence

will be the extent of intermarriage within the occupation. It is necessary therefore to discover the actual number of marriages that are occurring between the men and women engaged in the same occupations. A method must then be found to determine the number of marriages that would take place within the occupation were marriage merely a matter of chance. Finally, the number of marriages within the occupation must be compared with the results calculated as a matter of chance in order to show whether or not industry has been an appreciable factor.

It was decided that the only acceptable data would concern the occupation of the man and woman at the time of marriage.

An investigation of applications for marriage licenses in the various cities of the country showed a great mass of material that has never been digested. As Philadelphia is a typical city with a mixed industrial population and as the most complete information has been collected here, the investigation was limited to this city.

The data were derived from the books of the Marriage License Bureau in City Hall in Philadelphia. All licenses in the available books were used. They include practically all licenses issued in Philadelphia for the three years from June, 1913 to June, 1916. The numbers run from 299,758 to 332,586 and from 332,601 to 340,900 and from 341,401 to 355,367. The only exceptions made were those of Negroes, none of whom were counted. Throughout the work here and there a license was skipped because it related to a colored couple. This omission was made with the purpose of refining the statistics.

The grouping of occupations of the United States Census was used in a preliminary study of a thousand cases. Of these thousand cases 541 women worked and of these 275, more than half, married men in the same occupation. Of the thousand men more than 25 per cent. married into their own occupation.

In order to treat the matter at all, it was necessary to group the occupations. The groups selected were very carefully defined and the name of the group shows the type of person employed.

None	Unknown
Agriculture	Provisions
Clothing	Textiles
General Factory	Artisans
Drivers	Railroads and Street Railroads
Telegraph and Telephone	Finance
Retail Trade	Clerks
Clerical	Public Service
Professions	Semi-Professions
Personal Service	Domestic Service
Managers and Superintendents	Unskilled Labor
Firemen and Engineers	Printing
Miscellaneous	

Those in the class marked None are principally women living at home, but it also includes women who gave their occupation as housework. Under Unknown are four men and no women.

It is possible that more women are gainfully employed than these statistics show. A woman expecting to marry frequently gives her occupation as housework. This was a source of error. For example, the wife of a railroad president gave her occupation as housework. As she is a well-known society woman there could be no question of her being a domestic. For this reason it was decided to place all cases of housework in the group None. The result is that a great many domestics and some other women who were really at work are included among the group whose occupation is given as None.

Clerks are of two groups. The term clerk should not include salesmen. However, little effort is made in the Marriage License Bureau toward clear definition of occupation, and the result is that both salesmen and those belonging to the real clerical occupations are included in the group Clerks.

The desired information was coded into numbers according to a decimal system. The numbers used for the occupations were those standardized by the United States Census Bureau in their "Index of Occupations." These numbers and others representing the age and nationality of the couple and their parents were punched into tabulating cards to be sorted and counted by machinery. Records were secured of 49,000 cases; that is, the nationality, age and occupation of 98,000

brides and grooms were taken. The occupations of 98,000 fathers and the nationality of 294,000 individuals are also recorded.

The first separation of the cards was according to nationalities. Three groups were made. All cases where both bride and groom and the parents of both were born in the United States were put in one group and called the American Group. The Foreign Group was composed of the cases where all six persons were born outside the United States. All other cases were put in a group called the Mixed Group.

The material used in this investigation consists of the sworn answers to the questions on the application for a marriage license. All available applications made in Philadelphia from June, 1913 to June, 1916 were used except those of Negroes. Forty-nine thousand cases were collected which were sufficient to furnish two groups, one from each half of the cards, of five thousand American-born couples with all four parents of American birth. Each of these American groups is large enough to be typical of the whole number of American men and women of marriageable age in Philadelphia.

The results of most general interest have been condensed into the three following tables. Table I gives the occupations of the husbands and the wives in the total number of cases secured. Table II gives the occupations of the whole American Group and Table III for the part of the American Group taken from the first half of the cards. This group will be referred to as American Group No. 1. The statistics were refined to eliminate possible nationality factors for which reason the discussion will be confined to the American Group. But the results secured in the foreign and mixed groups are substantially the same, so that the statements made in regard to the American Group would apply also to these groups and perhaps would appear even stronger because of the larger numbers involved.

In the following tables the last column to the right gives the total number of women in each occupation. When read across the row, the tables show the occupations of the husbands. For example, in the first table the occupations of the husbands of the women without occupations may be found

by reading across the first row. In the fourth row across the total is 206, which is the number of women employed in the manufacture of provisions. The 9 in the fourth column indicates that 9 of these women married men also in the provisions group. The eighth column shows that 29 of them married men who were artisans.

A comparison of Tables II and III shows that in a great majority of cases the numbers in the complete American Group are double the corresponding numbers of the half-group No. 1. The frequency with which this occurs is an indication of two facts; in the first place, the tables are based upon sufficient numbers to be indicative of the proportion of people in each occupation in Philadelphia; secondly, any group of large size, taken at random, would fall in these same proportions. The American groups, correlated number for corresponding number throughout, give a Pearson coefficient of correlation of .9 plus. The diagonals, which represent the cases where like marries like, have a correlation of .97. This indicates a regular distribution dependent upon occupation. The significance of this high correlation is discussed later.

A study of these tables is helpful in the discussion which follows. In each table the diagonal from upper left to lower right show the numbers of like marrying like. It is apparent that each number on this diagonal represents a considerable proportion of the row and column in which it appears. The theory that occupation has a strong influence upon marriage selection would imply merely a disproportionate amount of marriage within certain allied occupations, an amount much less than has occurred. Moreover, there are sufficient marriages within the single occupational groups to make it noticeable. In many cases the number in the diagonal is the largest number in both row and column. That this should be so even in the small occupations is surprising since the chance expectation is that men in the large occupations will marry proportionately to the number of women in an industry whether the industry is large or small.

Were marriage merely a matter of chance, women in each occupation would marry men in occupations in proportion to the total number of men employed in each. Thus, if half of all

TABLE I. COMPLETE GROUP.

WOMEN / MEN	None	Unknown	Agriculture	Provisions	Clothing	Textiles	General Factory	Artisans	Drivers	Railroad and Street R. R.	Telephone and Telegraph	Finance	Retail Trade	Clerks	Clerical	Public Service	Professions	Semi-Professions	Personal Service	Domestic Service	Managers and Supts.	Unskilled Labor	Firemen and Engineers	Printing	Miscellaneous	Total
None	151	21	452	635	921	450	2893	3313	907	443	75	268	2427	1745	490	441	987	328	501	674	782	4195	479	263	303	24153
Unknown	1	—	.	.	.	.	1	.	.	.	.	.	2	.	.	.	.	.	.	.	.	.	.	.	.	4
Agriculture	.	.	2	.	2	.	.	5	.	.	.	.	1	1	.	.	.	1	2	.	3	.	1	.	.	18
Provisions	.	.	1	9	10	4	38	29	16	7	1	.	21	17	3	13	.	5	3	5	1	9	4	6	4	206
Clothing	9	5	58	102	585	115	679	944	205	100	30	24	637	394	118	77	106	88	108	96	131	171	67	113	57	4999
Textiles	1	1	7	11	43	180	299	294	74	21	2	7	62	80	26	21	9	27	7	24	16	142	29	20	19	1422
General Factory	9	2	22	70	205	103	666	603	213	52	19	5	288	167	57	77	26	35	46	61	31	306	46	78	34	3161
Artisans	.	.	1	1	1	2	10	8	2	2	1	1	4	1	2	2	.	.	.	1	.	2	.	2	1	44
Drivers	.	.	.	.	.	.	.	1	2	.	.	.	.	.	.	1	.	.	.	.	.	1	.	.	.	5
Railroad and Street Railway	.	.	1	.	.	.	.	.	.	1	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	2
Telephone and Telegraph	.	.	4	4	4	1	38	81	31	15	13	1	75	50	27	12	20	4	.	1	12	6	13	9	2	432
Finance	1	.	.	.	.	.	.	1	.	.	.	—	.	.	1	.	.	.	.	.	.	.	.	1	.	4
Retail Trade	1	1	22	33	38	22	162	233	46	58	6	8	384	146	57	43	36	41	18	21	62	19	21	28	15	152
Clerks	3	2	26	18	15	17	150	243	60	32	10	18	265	334	78	34	76	40	8	30	80	14	39	39	21	1642
Clerical	3	3	27	40	30	15	204	335	109	76	18	87	523	422	208	52	187	81	16	30	125	17	70	55	14	2747
Public Service	.	.	.	.	.	.	1	.	1	.	.	.	.	3	1	1	.	.	.	.	.	.	.	.	.	7
Professions	6	.	25	4	8	5	51	79	14	17	5	27	162	118	49	24	265	44	7	10	58	5	36	15	5	1039
Semi-Professions	7	2	26	6	9	5	55	90	29	22	3	6	91	59	20	19	89	97	7	23	43	12	26	8	12	766
Personal Service	1	1	2	7	6	2	27	30	9	6	1	1	29	18	7	6	11	7	11	25	11	14	3	.	2	237
Domestic Service	3	8	74	69	60	27	327	410	195	89	15	2	114	93	12	65	15	16	56	221	27	505	83	23	40	2559
Managers and Superintendents	.	.	1	.	.	.	4	5	2	1	2	.	4	3	3	.	2	.	1	.	5	.	5	1	.	39
Unskilled Labor	5	3	33	57	58	100	781	745	284	76	17	3	187	215	14	89	30	43	27	56	40	750	66	48	25	3758
Firemen and Engineers	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	.	1	.	1	—	.	.	2
Printing	1	.	4	4	8	4	50	55	17	9	1	1	29	26	3	5	5	2	.	8	3	11	7	35	3	291
Miscellaneous	.	.	1	1	6	8	14	16	3	2	.	.	7	10	2	3	1	2	.	4	6	7	47	3	6	149
Total	202	49	789	1071	1989	1069	6390	7520	2219	1029	219	459	5312	3901	1178	985	1865	861	818	1301	1436	6193	1042	747	563	49207

TABLE II. AMERICAN GROUP.

WOMEN / MEN	None	Unknown	Agriculture	Provisions	Clothing	Textiles	General Factory	Artisans	Drivers	Railroad and Street R. R.	Telephone and Telegraph	Finance	Retail Trade	Clerks	Clerical	Public Service	Professions	Semi-Professions	Personal Service	Domestic Service	Managers and Supts.	Unskilled Labor	Firemen and Engineers	Printing	Miscellaneous	Total
None	40	3	148	49	31	38	405	556	207	142	25	133	729	583	194	123	417	92	31	60	276	115	136	72	73	4678
Unknown	....	—	....	....	....	....	....	....	....	....	....	....	1	1	....	....	....	....	....	....	....	....	1	....	....	0
Agriculture	....	....	—	....	....	....	....	....	....	....	....	....	1	1	....	....	....	....	....	....	....	....	1	....	....	3
Provisions	....	....	....	5	....	3	9	6	5	3	....	....	5	5	1	4	....	1	....	1	1	2	2	....	1	54
Clothing	2	....	13	9	16	6	113	176	38	27	7	4	95	114	35	25	22	16	5	6	22	15	22	23	16	807
Textiles	....	....	1	3	1	27	52	66	21	8	2	6	5	23	6	10	3	2	1	4	5	8	2	5	8	269
General Factory	1	1	7	6	8	10	125	140	59	16	4	2	47	50	16	22	6	6	....	8	2	19	16	24	12	607
Artisans	....	....	....	1	....	....	4	8	1	1	1	....	2	....	....	1	....	....	....	1	....	....	....	....	....	15
Drivers	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	0
Railroad and Street Railway	....	....	....	....	....	....	....	....	....	1	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	1
Telephone and Telegraph	....	....	3	....	3	1	10	23	11	7	5	....	24	30	12	7	5	1	....	1	4	2	7	7	....	163
Finance	....	....	....	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	0
Retail Trade	....	....	12	5	5	1	36	65	16	20	3	6	66	64	17	14	12	12	1	7	7	5	6	8	6	388
Clerks	....	....	11	4	1	3	37	88	19	12	3	7	109	143	39	18	29	12	1	5	35	3	11	9	8	607
Clerical	1	....	10	12	6	3	61	124	29	32	10	13	195	173	73	11	72	33	3	10	61	2	18	15	7	973
Public Service	....	....	....	....	....	....	....	....	1	....	....	....	....	1	....	—	....	....	....	....	....	....	....	....	....	2
Professions	....	....	18	2	2	3	27	38	4	10	3	12	71	52	25	12	114	27	4	2	22	2	22	2	5	479
Semi-Professions	1	....	10	....	1	1	17	23	6	8	1	2	21	21	10	5	34	19	1	4	16	1	6	5	7	220
Personal Service	....	....	....	1	....	....	2	5	....	....	1	....	6	4	....	2	6	3	1	....	3	....	....	....	....	34
Domestic Service	1	....	8	9	4	2	26	48	43	11	2	....	24	18	2	11	1	5	2	32	5	28	14	1	5	302
Managers and Superintendents	....	....	....	....	....	....	1	1	....	....	....	....	1	1	2	....	....	....	....	....	3	....	3	1	....	13
Unskilled Labor	....	....	10	9	5	12	151	164	73	22	3	1	43	38	11	29	7	6	3	9	3	82	28	7	5	721
Firemen and Engineers	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	1	....	....	—	....	....	1
Printing	....	....	....	1	1	3	12	20	8	1	....	....	7	9	....	....	2	2	....	2	2	3	2	14	1	90
Miscellaneous	....	....	....	....	....	....	1	2	....	....	....	....	1	2	....	1	1	....	....	1	2	1	1	....	3	15
Total	46	4	251	116	84	113	1089	1548	541	321	70	180	1452	1332	442	275	731	237	53	154	469	288	297	193	156	10442

TABLE III. AMERICAN GROUP NO. 1.

MEN \ WOMEN	None	Unknown	Agriculture	Provisions	Clothing	Textiles	General Factory	Artisans	Drivers	Railroad and Street R. R.	Telephone and Telegraph	Finance	Retail Trade	Clerks	Clerical	Public Service	Professions	Semi-Professions	Personal Service	Domestic Service	Managers and Supts.	Unskilled Labor	Firemen and Engineers	Printing	Miscellaneous	Total
None	27	2	68	22	20	17	209	280	109	75	14	66	359	290	96	60	193	42	14	29	150	49	76	35	27	2329
Unknown	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	0
Agriculture	....	....	—	....	....	....	....	....	....	....	....	....	1	1	....	....	....	....	....	....	....	....	1	....	....	3
Provisions	....	....	....	3	....	1	5	2	4	2	....	....	4	4	....	2	....	....	....	1	....	1	1	....	....	30
Clothing	1	....	5	6	10	2	53	85	16	12	4	3	50	60	21	16	14	8	1	3	13	9	13	15	6	426
Textiles	....	....	1	2	....	7	26	34	11	5	....	1	5	12	5	5	....	....	....	1	3	4	1	2	5	130
General Factory	1	....	4	4	2	6	58	69	24	8	3	1	28	20	11	16	2	2	....	6	1	8	8	8	4	294
Artisans	....	....	....	1	....	....	2	1	....	1	1	....	2	....	....	....	....	....	....	....	....	....	....	....	....	8
Drivers	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	0
Railroad and Street Railway	....	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	0
Telephone and Telegraph	....	....	2	....	3	1	7	16	8	3	2	....	14	13	7	3	2	1	....	....	....	1	3	1	....	87
Finance	....	....	....	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	0
Retail Trade	....	....	4	3	1	....	16	28	9	13	....	1	37	35	10	10	6	4	....	1	4	3	4	3	3	195
Clerks	....	....	3	1	....	1	21	45	8	4	....	4	57	80	28	8	16	2	1	5	11	....	6	5	6	312
Clerical	....	....	4	5	2	3	24	55	16	17	6	5	106	96	39	4	37	21	2	2	30	....	8	9	3	494
Public Service	....	....	....	....	....	....	....	....	....	....	....	....	1	....	....	—	....	....	....	....	....	....	....	....	....	1
Professions	....	....	12	....	1	1	9	17	2	3	....	8	33	26	17	4	66	10	4	1	18	2	11	1	2	248
Semi-Professions	1	....	4	....	1	1	7	13	3	6	1	....	13	7	7	4	18	6	1	2	8	1	4	....	2	110
Personal Service	....	....	....	....	....	....	1	4	....	....	1	....	2	2	....	2	4	2	1	....	1	....	....	....	....	20
Domestic Service	....	....	6	8	3	2	13	25	21	1	1	....	16	10	2	4	....	1	2	18	2	11	7	1	4	158
Managers and Superintendents	....	....	....	....	....	....	1	....	....	....	....	....	....	1	1	....	....	....	....	....	3	....	2	....	....	....	8
Unskilled Labor	....	....	3	6	2	2	81	84	35	11	2	1	31	23	5	18	4	3	1	6	....	46	10	5	2	375
Firemen and Engineers	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	—	....	....	0
Printing	....	....	....	1	1	....	7	11	1	1	....	....	2	7	....	....	2	2	....	1	2	....	1	3	....	48
Miscellaneous	....	....	....	....	....	....	1	1	....	....	....	....	1	2	....	1	1	....	....	....	2	3	....	....	1	13
Total	30	2	116	62	46	44	541	770	267	162	35	90	761	690	249	157	365	104	27	76	248	132	156	94	65	5289

men were employed in manufactures, it would be found that one half of the manufacturing women, one half of the professional women, one half of the women clerks, etc., would marry men in manufactures.

If a girl enters a profession, there is one probability in four that she will marry a professional man and small likelihood that she will marry into manufacturing. Ability to predict the occupation of the man whom a woman will marry may well be regarded as belonging in the realm of clairvoyance. Yet, while it will always be impossible to prophesy for an individual woman, for a group of women such prediction is not only easy but extremely accurate.

For this work it was necessary to secure a coefficient of propinquity-expectation. This is the ratio between the number of cases that would be predicted by the laws of chance and the number that actually occur.

The following illustration will exemplify the laws of chance. Place in a bag thirty-one balls and thirty-one cubes, each of which is either red, blue or green. Shake the bag and then draw at random the balls and cubes in pairs, one ball and one cube at each draw. If this process is repeated many times, it will be found that of the possible combinations of colors certain ones occur more frequently, depending upon the number of each color. If there are twice as many red balls and red blocks as of any other color, there will be four times as many chances of drawing a red pair as of drawing a pair of any other color.

If however, the red balls and cubes are placed in one part of the bag, the blue balls and cubes in another part of the bag, the green balls and cubes in a third part and the bag shaken very slightly, the chances of drawing two of a color are considerably increased. New tables can be constructed showing the effect of propinquity. The drawing is still made at random, by chance, and yet there is a decided increase of possibility that like will be paired with like.

Were marrying merely a matter of chance, the first illustration would explain the relation of occupation to marriage selection. The balls and cubes represent men and women respectively; the colors symbolize their occupations. The

amount of intermarriage would be directly proportional to the numbers employed in each line of work. However, the second illustration typifies the actual influence of occupation. The position of the balls and cubes in the bag corresponds to occupational propinquity in life. Women in the home have no occupational propinquity and they marry entirely by chance as the blocks were drawn in the first illustration. The number of their husbands from any industry is in direct proportion to the number of men employed in that industry. This is shown in the tables. However, they are the only group of women thus subject to chance and the tables show that they are the only group of women who have married husbands in these proportions. One tenth of the total number of men are employed in general factory work. Yet in some cases more, in others less, than one tenth of the women in an occupation marry men in general factory work, the proportions varying according to the particular occupation of the wives.

Questions of method arise, first of determining the expectation of marriage between people of various occupations and second of comparing this expectation with the empirical data. A small table will serve as an illustration. Suppose that there are three possible occupations and that men and women are distributed among these occupations. The table should be read across for women and down the columns for men.

EMPIRICAL TABLE.

	A	B	C	Totals
A.....	**5**	8	7	20
B.....	3	**3**	1	7
C.....	1	1	**2**	4
Totals	9	12	10	31

There are 9 men and 20 women in the first occupation, 12 men and 7 women in the second and 10 men and 4 women in the third. This would give a total of 31 men and 31 women. A marriage would be possible between any of these men and each of these women; that is, there would be 31 possible marriages for each woman and for each man. Since there are 20 women and 9 men in Group A, there are 9 times 20 or 180

possible combinations between the men and women in Group A. Likewise, the number of possible combinations between the men in Group A and the women in Group B is 9 times 7 or 63 and between men in Group B and women in Group C it is 48. If a table of possible combinations is made, it will be as follows:

	A	B	C	Totals
A.....	**180**	240	200	620
B.....	63	**84**	70	217
C.....	36	48	**40**	124
Totals.....	279	372	310	961

There are 961 possible combinations but only 31 marriages are possible. Therefore, if each number of the above table is divided by 31, a table will be formed which will give what could be expected to occur were marriage merely a matter of chance. This Expectation Table will show the distribution of the 31 marriages among the people in the three occupation groups. The distribution of marriages represented in this table corresponds to the distribution of colors when the balls were drawn at random.

EXPECTATION TABLE.

	A	B	C	Totals
A.....	**5.8**	7.7	6.4	20
B.....	2.0	**2.7**	2.3	7
C.....	1.2	1.5	**1.3**	4
Totals.....	9	12	10	31

A comparison of the Empirical Table with the Table of Expectation shows where actuality exceeds the expectation and where it is less than expectation. The diagonal contains all cases of like marrying like and indicates that Group A married Group A less than would have been predicted by chance while in the other two cases the reality slightly exceeded expectation.

Dividing each number in the Empirical Table by the corresponding number in the Table of Expectation gives a third

TABLE IV. EXPECTATION TABLE FOR AMERICAN GROUP.

	None	Unknown	Agriculture	Provisions	Clothing	Textiles	General Factory	Artisans	Drivers	Railroad and Street Railway	Telephone and Telegraph	Finance	Retail Trade	Clerks	Clerical	Public Service	Professions	Semi-Professions	Personal Service	Domestic Service	Managers and Superintendents	Unskilled Labor	Firemen and Engineers	Printing	Miscellaneous
ge	20.61	1.79	112.4	51.97	37.63	50.62	487.9	693.5	242.4	143.8	31.36	80.64	650.5	596.7	198.02	123.2	327.5	106.2	23.74	68.99	210.1	129.02	133.06	86.46	69.9
known	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
iculture	.013	.0011	.672	.033	.024	.032	.313	.445	.155	.092	.0201	.052	.417	.383	.127	.079	.210	.068	.015	.044	.135	.083	.085	.055	.045
visions	.238	.0207	1.30	.800	.434	.584	5.63	8.01	2.80	1.66	.362	.931	7.51	6.89	2.29	1.42	3.78	1.23	.274	.796	2.43	1.49	1.54	.998	.807
thing	3.56	.309	19.40	8.96	6.49	8.73	84.16	119.6	41.81	24.81	5.41	13.91	112.2	102.9	34.16	21.25	56.49	18.32	4.096	11.902	36.25	22.26	22.95	14.92	12.056
iles	1.19	.103	6.47	2.99	2.16	2.91	28.05	39.88	13.94	8.27	1.803	4.64	37.41	34.31	11.39	7.084	18.83	6.11	1.37	3.97	12.082	7.42	7.65	4.97	4.019
eral Factory	2.67	.233	14.59	6.74	4.88	6.57	63.30	89.99	31.45	18.66	4.069	10.46	84.41	77.43	25.69	15.99	42.49	13.78	3.081	8.95	27.26	16.74	17.26	11.22	9.068
ans	.066	.0057	.361	.167	.121	.162	1.56	2.22	.777	.461	.101	.259	2.086	1.91	.635	.395	1.05	.3405	.076	.221	.674	.414	.427	.277	.224
ers	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
R. and St. R. R.	.0044	.00038	.024	.011	.0080	.011	.104	.148	.052	.031	.0067	.0172	.139	.128	.042	.026	.070	.023	.0051	.015	.045	.028	.028	.018	.015
phone and Telegraph	.718	.062	3.92	1.81	1.31	1.76	17.00	24.16	8.45	5.01	1.69	2.81	22.67	20.79	6.90	4.29	11.41	3.70	.827	2.404	7.32	4.50	4.64	3.013	2.44
ance	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
ail Trade	1.71	.149	9.33	4.31	3.12	4.20	40.46	57.52	20.10	11.93	2.601	6.69	58.95	49.49	16.42	10.22	27.16	8.81	1.97	5.72	17.43	10.701	11.036	7.17	5.80
rks	2.67	.233	14.59	6.74	4.88	6.57	63.30	89.99	31.45	18.66	4.069	10.46	84.41	77.43	25.69	15.99	42.49	13.78	3.081	8.95	27.26	16.74	17.26	11.22	9.068
rical	4.29	.373	23.39	10.81	7.83	10.53	101.47	144.2	50.41	29.91	6.52	16.77	135.3	124.1	41.19	25.62	68.12	22.08	4.94	14.35	43.702	26.84	27.67	17.98	14.54
lic Service	.0088	.00077	.048	.022	.016	.022	.209	.296	.104	.061	.013	.034	.278	.255	.085	.053	.140	.045	.0102	.029	.090	.055	.057	.037	.030
essions	2.11	.183	11.51	5.32	3.85	5.18	49.96	71.01	24.82	14.73	3.21	8.26	66.61	61.102	20.28	12.61	33.53	10.87	2.43	7.064	21.51	13.21	13.62	8.85	7.16
si-Professions	.969	.084	5.29	2.44	1.77	2.38	22.94	32.61	11.40	6.76	1.47	3.79	30.59	28.064	9.31	5.79	15.40	4.99	1.12	3.24	9.88	6.068	6.26	4.066	3.29
onal Service	.150	.013	.817	.378	.274	.368	3.55	5.04	1.76	1.05	.228	.586	4.73	4.34	1.44	.895	2.38	.772	.173	.501	1.53	.938	.967	.628	.508
mestic Service	1.33	.116	7.26	3.35	2.43	3.27	31.50	44.77	15.65	9.28	2.025	5.21	41.99	38.52	12.78	7.95	21.14	6.85	1.53	4.45	13.56	8.33	8.59	5.58	4.51
rs. and Supts.	.057	.0050	.312	.144	.105	.141	1.36	1.93	.674	.400	.087	.224	1.81	1.66	.550	.342	.910	.295	.066	.192	.584	.359	.370	.2403	.194
killed Labor	3.18	.276	17.33	8.01	5.80	7.802	75.19	106.9	37.36	22.16	4.83	12.43	100.3	91.97	30.52	18.99	50.47	16.36	3.66	10.63	32.38	19.89	20.51	13.33	10.77
men and Engineers	.0044	.00038	.024	.011	.0080	.011	.104	.148	.052	.031	.0067	0172	.139	.128	.042	.026	.070	.023	.0051	.015	.045	.028	.028	.018	.015
nting	.396	.034	2.16	1.000	.724	.974	9.39	13.34	4.66	2.77	.603	1.55	12.51	11.48	3.81	2.37	6.30	2.043	.457	1.33	4.042	2.48	2.56	1.66	1.34
ellaneous	.066	.0057	.361	.167	.121	.162	1.56	2.22	.777	.461	.101	.259	2.086	1.91	.635	.395	1.05	.340	.076	.221	.674	.414	.427	.277	.224

table, of which each number is the ratio between the actuality and the expectation. This may be called the coefficient of expectation or in this case the coefficient of propinquity expectation.

COEFFICIENT TABLE.

	A	B	C
A.....	**.86**	1.04	1.09
B.....	1.50	**1.11**	.43
C.....	.83	.67	**1.54**

The diagonal further indicates that in the first group the number of cases where like married like is less than the expectation. In the second and the third group the index of measurement is more than unity. Where the index exceeds unity, the reality is greater than the expectation. Where the index is less than unity, expectation is greater than reality. The total chances of like marrying like may be determined by dividing the sum of the diagonal of the Empirical Table by the sum of the diagonal of the Expectation Table. In this case like marries like slightly more than the expected number of times.

By the method illustrated the following table of expectation, Table IV, was computed for the American Group. A comparison of this table with Table II shows that each number in the diagonal of the Empirical Table exceeds the corresponding number in the diagonal of the Table of Expectation; that is, like marries like beyond expectation. If marriage were purely a matter of chance as far as occupation is concerned, Table II and Table IV would be very nearly identical. However, there are but few places where they approach identity.

Table V is a detailed comparison in the American Group of the empirical table of marriage, Table II, with the expectation table of pure chance, Table IV. Each number in this table is a coefficient of expectation, that is, it is the ratio between the number of times a marriage occurred and the number of times that it might have been predicted by chance. Numbers greater than unity represent more intensive intermarriage and numbers less than unity represent combinations that occur less frequently than expectation. This table is the main center of interest in the statistical presentation of the material. The

TABLE V. TABLE OF COEFFICIENTS FOR AMERICAN GROUP.

	None	Unknown	Agriculture	Provisions	Clothing	Textiles	General Factory	Artisans	Drivers	Railroad and Street Railway	Telephone and Telegraph	Finance	Retail Trade	Clerks	Clerical	Public Service	Professions	Semi-Professions	Personal Service	Domestic Service	Managers and Superintendents	Unskilled Labor	Firemen and Engineers	Printing	Miscellaneous
None	1.94	1.68	1.32	.943	.824	.751	.830	.802	.854	.987	.797	1.65	1.12	.977	.980	.998	1.27	.866	1.31	.870	1.31	.891	1.02	.833	1.04
Unknown	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Agriculture	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Provisions	....	....	....	2.33	....	5.14	1.60	.749	1.79	1.81	....	....	.666	.726	.437	2.82	....	....	....	....	.412	1.34	1.30	....	....
Clothing	....	....	.670	1.00	2.47	.687	1.34	1.47	.909	1.09	1.29	.288	.847	1.11	1.02	.235	.389	.873	1.22	.504	.607	.674	.959	1.54	1.33
Textiles	....	....	.155	1.00	.463	2.28	1.85	1.65	1.51	.967	1.11	1.29	.134	.670	.527	1.41	.159	.327	.730	1.01	.414	1.08	.261	1.01	1.99
General Factory	....	....	.480	.890	1.64	1.52	2.06	1.56	1.88	.857	.983	.191	.557	.646	.623	1.38	.141	.435	....	.894	.073	1.14	.927	2.14	1.32
Artisans	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Drivers	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
R. R. and St. R. R.	....	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....
Telephone and Telegraph	....	....	....	....	....	....	.588	.952	1.30	1.40	4.59	....	1.06	1.44	1.74	1.63	.438	.270	....	.416	.546	.444	1.51	2.32	....
Finance	....	....	....	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....	....	....	....	....
Retail Trade	....	....	1.29	1.16	1.60	.238	.890	1.13	.796	1.68	1.15	.811	1.22	1.29	1.04	1.37	.442	1.36	.508	1.22	.402	.467	.544	1.12	1.03
Clerks	....	....	.754	.593	.205	.457	.585	.978	.604	.643	.737	.669	1.29	1.85	1.52	1.13	.683	.871	.325	.559	1.28	.179	.637	.802	.882
Clerical	....	....	.428	1.11	.766	.285	.601	.860	.575	1.07	1.53	.775	1.44	1.39	1.75	.429	1.06	1.49	.607	.697	1.40	.075	.651	.834	.481
Public Service	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	—	....	....	....	....	....	....	....	....	....
Professions	....	....	1.56	.376	.519	.579	.540	.535	.161	.679	.935	1.45	1.07	.851	1.23	9.52	3.40	2.48	1.65	.283	1.02	.151	1.62	.226	.698
Semi-Professions	....	....	1.89	....	.565	.420	.741	.705	.526	1.18	.680	.528	.686	.748	1.07	.864	2.21	3.81	.893	1.23	1.62	.165	.958	1.23	2.13
Personal Service	....	....	....	....	....	....	.563	.992	....	....	....	....	1.27	.922	....	2.23	2.52	3.89	5.78	....	1.96	....	....	....	....
Domestic Service	....	....	1.10	2.69	1.65	.612	.825	1.07	2.75	1.19	.988	....	.572	.467	.156	1.38	.047	.730	1.31	7.19	.369	3.36	1.63	.179	1.11
Mgrs. and Supts.	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	5.14	....	8.11	....	....
Unskilled Labor	....	....	.577	1.12	.862	1.54	2.01	1.53	1.95	.993	.621	.080	.429	.413	.360	1.53	.139	.367	.820	.847	.093	4.13	1.37	5.25	.464
Firemen and Engineers	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	66.67	....	....	—	....	....
Printing	....	....	....	....	....	3.08	1.28	1.50	1.72	.361	....	....	.560	.784	....	....	.317	.979	....	1.50	.495	1.21	.781	8.43	.746
Miscellaneous	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	....	—

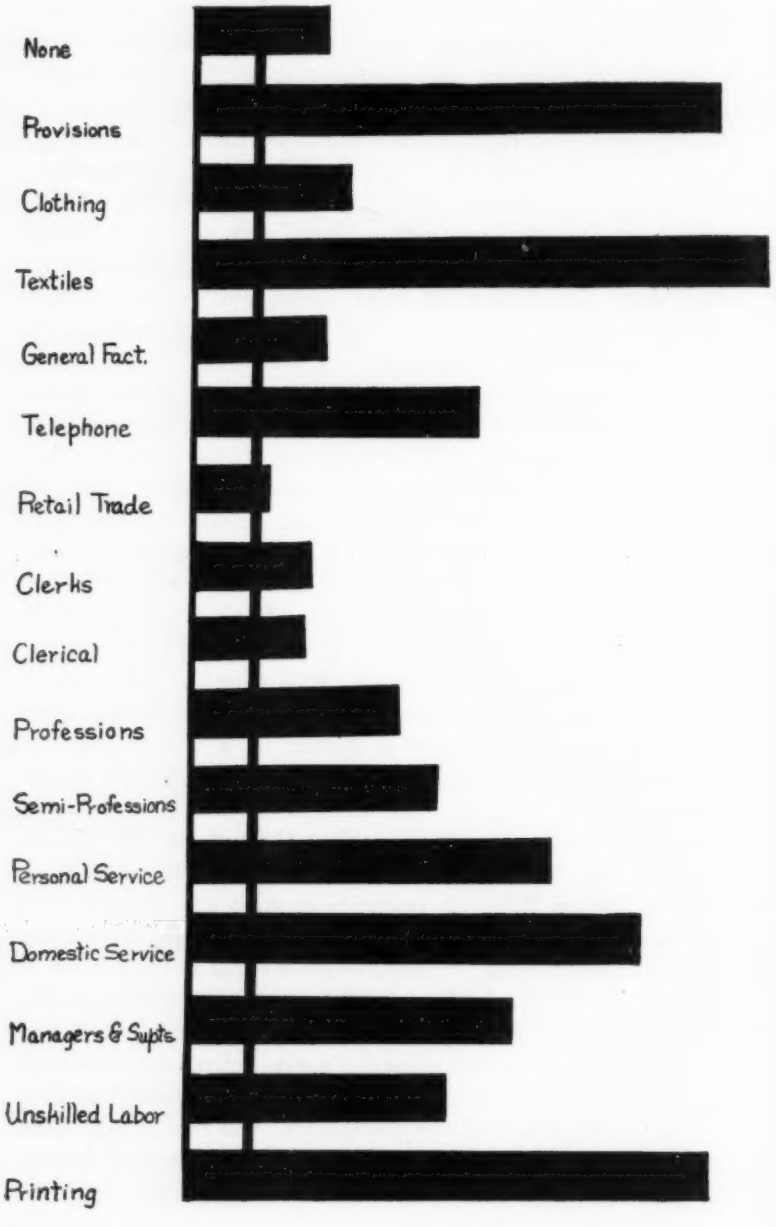
diagonal represents the coefficients of marriage within each occupation. Each number in the diagonal exceeds unity and for most of the occupations the ratio is greater than two. Because of the interest and importance of marriage within the occupation, a graph of this diagonal was made.

This graph represents the ratios of like occupations marrying like. The solid black line represents unity of expectation. The columns represent the ratio of the actuality to this unit of expectation. In all cases where there were sufficient numbers to permit of statistical treatment, the reality exceeds the expectation. In all but the least skilled and most undifferentiated occupations there is at least twice the expectation and many times this is far exceeded.

When the numbers in the diagonal are added, the sum gives the number of cases of like occupations marrying like. When the sum of the diagonal of the Empirical Table is divided by that of the Table of Expectation the quotient is 2.79. For the mixed group it is 2.50 and for the foreign group it falls to 2.00. These are the actual coefficients of propinquity expectation for large groups in Philadelphia and it is very probable that these figures are approximated in most of the industrial cities where similar conditions exist. This means that among Americans of American parentage, 2.79 times as many marriages are occurring within the occupations as would take place were marriage merely a matter of chance. For the mixed group it is 2.5 times as many and for the foreign group twice as many.

The statement has been made that throughout the American Group the coefficient of expectation for like occupation marrying like is 2.75. A comparison with the results of a research made in connection with this throws an interesting light on this coefficient. Using the same cards as the source of material, a study was made of the intermarriage among nationalities. Tables of expectation and of coefficients were made The coefficients for intermarriage of natives of native parentage was found to be 1.8 while that of intra-occupational marriage is 2.75. Irrespective of the occupation, the rate of marriage within it is greater than between sons and daughters of American parents. There is twice as much marriage between

GRAPH.—SHOWING RATIO OF LIKE MARRYING LIKE.



employees of textile mills as between the children of German-born parents. The important deduction is not that there is so little intermarriage within lines of nationality but that there is so much intermarriage within occupations.

Any ambiguity of results given so far is probably due to their applying to intermarriage of groups of people rather than to those within a particular occupation. The fact that there are hundreds of subdivisions of occupation in the textile mills, in the boot and shoe industry, etc., made statistical treatment of particular occupations impossible. Another study was made in order to determine to how great an extent propinquity is a matter of the particular occupation.

A supplementary investigation was made of statistics of Bryn Mawr alumnae. It was found that about 90 per cent. of the married alumnae had married college graduates. More than 60 per cent. of them married men in professions. Out of ten doctors who graduated from Bryn Mawr and later married, seven married doctors. There was a great deal of intermarriage within such groups as school teachers, social workers, doctors and musicians. These figures were obtained through the courtesy of Dean Isabelle Madison from a list of the occupations of over five hundred alumnae and their husbands. This list contained all cases that could be collected.

The statistical results did not differ greatly from those given in this paper. Yet here was a group of women who had been trained along highly specialized lines. Here, if anywhere, would be a chance to work out results significant for their individual and human meaning rather than the statistical. The grouping of occupations in this paper was owing to the necessity for groups large enough to give the results a statistical meaning. The Bryn Mawr study gave an opportunity for interesting individual treatment. Cases of teachers, doctors and others easily classified gave results similar to those obtained in the large groups already discussed, except that there was a greater degree of intra-occupational marriage. Thirty cases of odd or unusual occupations difficult to classify remained.

Of these thirty marriages there were sixteen in which there was no similarity of occupation but the other fourteen married as follows:

Women.	*Men.*
Explorer	Explorer
Lawyer	Lawyer
Farmer	Farmer
Manager of Photo Plays	Manager of Photo Plays
Art Student	Landscape Engineer
Editor	Editor
College Secretary	College Secretary
Newspaper Writer	Reporter
Law-office Secretary	Law Student
Assistant Superintendent of Children's Bureau	Secretary of Children's Bureau
Settlement Worker	M. D. Settlement Worker
Student of Sociology	Student of Sociology
Student of Architecture	Civil Engineer
Student of Biology	Professor of Biology

These cases give strong evidence of the influence of occupational propinquity since it would be impossible for all these unusual marriages to occur merely as a matter of chance.

Suitable statistical evidence as to the sufficiency of the data comes from two mathematical comparisons. The first of these consists of a comparison of the number of individuals in each of the occupations of this complete group with the total number in each occupation in Philadelphia. This is found to be nearly a simple proportion throughout.

The second source of evidence is even more convincing because it concerns only those of marriageable age in Philadelphia. It is a refined grouping. Both of the American groups were selected at random in a statistical manner—by pure chance. Thus, each is a statistically typical group. Yet the number in each occupation in one group correlated with the corresponding number in the other group gave a coefficient of correlation of .9 plus. It is upon such evidence as this that the ratio of the number of men in a particular occupation to the total number of men in the group is considered as a fair index of the theoretical possibility of marriage in each occupation.

The proof as to the constant distribution of data consists of two parts. In the first place, great emphasis has been laid

upon the amount of intermarriage within the occupation. This is represented by black-faced diagonals in the tables. The Pearson coefficient of correlation of the diagonals in the two American groups is .97. This indicates that at the time shown there was a constant effect of propinquity throughout the American Group. This high correlation held in the other tables. The second point of interest was whether or not the tables were sufficient to show the general distribution of marriage so that it was fair to predict the chance of marriage on this basis. This would be shown by the correlation, number by number of the two American groups. It is impossible to determine by inspection. However, the Pearson coefficient of correlation for each number of one table correlated with the corresponding number of the other table was worked out and found to be .9 plus.

Study of occupational propinquity upon marriage selection leads to the conclusion that a girl's choice of her vocation has a directing influence in deciding from which occupation she will marry her husband. Tables giving the proportion of fathers and sons and of fathers and daughters in the same occupation were made. They give new results. Computation of the coefficients of expectation shows that the chances are even for a man to marry into his own occupation and for his son to enter the same occupation. The close relationship between the occupations of father and son has long been a well-established fact but woman in industry is so recent a phenomenon that there has been no realization of the close occupational relationship of husband and wife. These tables with other similar data will make possible a new type of statistics. With almost as much certainty as is offered by the Mendelian Law it will be possible to predict what the chances are for a girl entering a particular occupation to have her children enter the same occupation. These results studied with reference to nationality will demonstrate the process of climbing the ladder of industries supposed to be typical of the development of a race. In order to obtain such specific deductions, it is necessary to deal with large numbers but as the movement already involves one half of the marrying women in the industrial cities of the country, it is large enough to merit consideration.

In modern industry where the large numbers of women and the selection of abilities by different occupations exist, *specialization of ability through heredity, environment, and training is favored by the propinquity of like types of men and women in industry.* To whatever extent an industry requires specialized ability, it may be bred by the matings within the industry. As the conditions of an occupation change, a new type of ability will be required and the breeding will mechanically change to the new type. No psychological examination to determine the complex required by the industry is needed for this breeding—survival within the industry is the test. The three biological essentials are present, heredity, training, and environment. Where men and women are employed together in like positions, each profession or specialized occupation is the natural breeding place for people of the type of ability required. This situation must be recognized both by those who believe in heredity and by those who favor environment. Sex propinquity in modern industry seems destined to affect the matings and through the matings the type of the coming generations.

NOMINATIONS FOR OFFICERS OF THE ASSOCIATION 1918–1919.

By virtue of appointment by President Wesley C. Mitchell, and to carry out the provisions of Article 9 of the Constitution of the Association, we submit herewith nominations for officers of the Association for the year 1918–1919 as follows:

President

JOSEPH A. HILL

Vice-Presidents

WILLARD C. BRINTON
CARROLL W. DOTEN
WILFORD I. KING

Counsellors

WESLEY C. MITCHELL
FREDERICK L. HOFFMAN
WALTER F. WILCOX

Treasurer

S. B. PEARMAIN

Secretary

ROBERT E. CHADDOCK

Librarian

HORACE L. WHEELER

Editor

JOHN KOREN

Signed: WILLIAM A. HATHAWAY,
FREDERICK S. CRUM,
ALLYN A. YOUNG,
Nominating Committee.

REVIEWS AND NOTES.

Introduction to Mathematical Statistics. By Carl J. West, Ph.D. Columbus, Ohio: R. G. Adams and Co., 1918. Pp. 144.

This book affords a mixed feeling of delight and disappointment.

According to his preface, the author addresses himself to statisticians who, not being also mathematicians, are unacquainted, or only partly acquainted, with the theory underlying many of the statistical processes used by them, or that might be profitably used if they could do so understandingly. To avoid the limitation that has made the existing treatises on the subject unserviceable to these persons, he aspires to convey the matter in more elementary mathematical media. It is not a handbook on statistics, for it assumes a familiarity with the ordinary statistical methods. Rather, the author seeks, not to "popularize," but to make accessible to larger numbers a knowledge of the science underlying an art that is practiced, in many respects, on the basis merely of authority.

The reviewer's response to the note in this preface was instant and exultant, for he is himself one of the large class addressed. He dares to believe, moreover, that this feeling and his subsequent experience with the book were somewhat representative, and therefore ventures to set down, for the sake of the common good, his impressions after reading it.

We are correctly described by the author as frequently uninformed on the reasons for our statistical conduct. While we earnestly desire to be instructed, we are handicapped by the limits upon our knowledge of higher mathematics. Perhaps the reviewer's maximum attainment, comprising introductory courses through analytical geometry and calculus, is typical. Perhaps, also, his present vacant memory of the content of those courses, long since completed, in an unpracticed subject is even more typical. The available books that we have boldly opened have proven quite beyond us, except at most in spots, and those near the beginning. We are troubled with the fidelity of their authors to the technical terminology and forms of expression of mathematics. We are suffocated with the mathematical atmosphere. We are exhausted by the mathematical gymnastics they put us through. We are utterly baffled by the long leaps from "since so-and-so is true, it is apparent that such-and-such follows," when it is not only not apparent to us, but not at all fathomable even after hours of soundings with our inadequate plumb line. In a word, our minds are non-mathematical. We may even be called boneheads by the plain spoken. What we need is a presentation of the theory of our art in terms boldly defying the standards of good form for mathematical expression, and without mathematical exercises thrown in for our mental discipline. We would rather read ten pages of utterly naive explanation that instantly carries a point to our undeveloped mathematical comprehension than spend ten times as long on one page of scrupulously scientific and coldly dignified stuff that

never touches us. Speaking etymologically, we would like a book that will patiently *instruct* us and not aim to *educate* our mathematical abilities or presuppose any mental agility with the concepts and terminology of higher mathematics.

This, then, is the group that the author has essayed to befriend. It is from their point of view, and not that of a mathematician, that this review is written.

The general plan of the book lends itself to an easily comprehended development. The author begins with the familiar graphical device of a wavy line connecting successive points and gives this a simple mathematical expression as a "curve" located with respect to its axes and amply equipped with definitions. He describes the significance of such a curve and of the area within it, and the theoretical meaning of a smoothed curve. From this the nature of the frequency curve is developed in mathematical terms that facilitate later discussion, and its aptitude for interpreting typical, although incomplete, data is explained. A chapter on averages then describes the mean, both arithmetical and geometric, the median and the mode, and explains the peculiar statistical properties and serviceableness of each. Then follows an analysis of the form of a frequency curve as disclosing the nature of any distribution and the dispersion of the data. Measures of dispersion, their mathematical significance and their comparative virtues are explained. A chapter is given to the normal frequency curve, in which the theory of it is stated and applied in fitting such a curve to given data.

Then follow chapters on the correlation of series of data. In these the author's plan of discussion is most successful in its adaptation to the readers whom he addresses. He avoids the too mathematical terminology and concepts employed in the explanations of the theory of correlation commonly found. He builds onto the concepts of the frequency curve, already made familiar, and develops by simple steps the nature of correlation, the evidences of its existence, and the measure of its degree. The subject is given further special treatment in chapters on correlation from ranks and on the method of contingency.

A chapter on the moments of a distribution explains how the equation may be found for a straight line to fit data susceptible of being so interpreted. In an appendix, the fitting of frequency curves to statistical distributions as developed by Pearson is explained and formulas given for those readers who are more adept in mathematics than the class to which the author mainly addresses himself.

The delightful features of the book lie in the simply planned selections of essential topics and in the evident attempt to develop them in language understandable by ordinary statisticians. With perseverance, the reader of the class addressed will be pleased with the grasp he will get of the theory of statistics as presented in this volume.

But this delight, although great, will be marred by disappointments. In the first place, while well planned, the discussion is not well composed. It is often obscurely phrased. The point of departure is in nearly every

case that of the mathematician, discursing on abstract mathematical formulae, rather than that of the concrete statistical setting in which the statistician finds his problem. The attention of a non-mathematical statistician would be far more likely secured, his interest better sustained, and his comprehension facilitated and quickened if the chapters began with a concrete statistical problem for which the theory was then developed as the method of solution. Not only are the mathematical topics introduced without previous reference to their statistical bearing, but they are abruptly introduced. There is want of introductory and transitional passages to hook up the matter under discussion with what had preceded, as well as with concrete settings. The effect is to leave the reader with insufficient perception of what it is all about and how it bears on what he already knows of statistical methods.

The author's expression further misses the mark at times in that, notwithstanding the limitations on his selected group of readers, he too often writes as a mathematician, assuming that terms, condensed or elliptical expressions, and suggested explanations that would be quite obvious to mathematicians will be clear to those whom he addresses. While the statisticians addressed may, indeed, worry out the meaning, they will be irritated, if not discouraged, at having to work out puzzles before they can get at the matter that interests them. The whole book extends to only 144 pages of text. It could well afford to have been doubled in volume in order to include fuller definitions and greater statement of detail.

Another source of confusion is a number of inaccuracies. Many of these are plainly typographical, such as inevitably occur in setting up mathematical copy. Some of these are obvious to any reader. For example the omission of the dollar sign from the mean prices found on page 33; the omission of "q" from "frequency," on page 28; and the transposition of the side and top legends in Table IX, page 69. Another less obvious, and hence more perplexing error is found on page 40, where the paragraph at the middle of the page beginning "The median can," etc., and Exercises 19 and 20 apparently should come at the top of the page. These defects will give readers little or no trouble. But thoughtful proofreading ought not to have permitted such perplexing errors as the following: the substitution of the digit 2 for 5 in the footing of the fourth column on page 50, which does not agree with the totals for the same items as shown on pages 24 and 33; the insertion of the subscript 3 with the parenthetical term near the bottom of page 87; the substitution of the figure 6 for the letter y in the expression for S_3, and the ommission of the subscript 1 from the letter y in the expression for S_5, on page 98; the substitution of the exponent 2 for 3 in the term d in the reduction of moments on page 99; the reference to page 91 instead of 96 for formulas used on page 101; the substitution of the subscript y for x with $\bar{y}$ in the equation on page 108.

Of like disconcerting effect is occasional looseness, if not inaccuracy, of expression. On page 48, the paragraph defining the mean squared deviation is headed "The Standard Deviation," while the definition of the standard deviation is given on page 51. In the text on page 49 the Greek letter

sigma is designated to represent the mean squared deviation, while in the accompanying equation, and in subsequent usage it is designated by the usual sigma squared. On page 83, the mean squared deviation of the regression means is derived and is then divided, according to the text, by the *standard deviation* of all the corresponding array, but according to the equation, by its *mean squared deviation*. At the bottom of page 112, reference is made to the mean standard deviation of certain quantities when the mean squared deviation is meant.

In working out the computations for the formulas at the bottom of page 106, the reviewer has not been able to derive the same solutions as those shown on page 107.

Another apparent inaccuracy is so consistently followed as to leave the reviewer in doubt whether the author has made a mistake or is himself simply dense. In applying the definition, on page 71, of the respective means of the x- and y- arrays to the problem as described on pages 75 and 76, and tabulated on page 77, it has seemed that the x- array has been confused with the y- array. On page 82 the values shown for the standard deviations of x and y seem to be a transposition of the values required by the definitions near the head of the page. This is essentially the same difficulty as that just noted. It is present also in the interpretation of the example on pages 78 and 79, for the reviewer cannot follow the explanation of the second distribution unless he takes for the y's what seems, by definition to be the x's. The probability is against the author repeatedly making the same mistake. But, then, he is guilty of most indefinitive definitions and non-explanatory explanations for the readers he hopes to assist. The reviewer intends, for his own enlightenment, to submit this to someone more qualified than he to pass upon the mathematics of the case. But he here records, as one of the readers addressed, the results which the book itself has produced on this point.

The effect of the author's, at times too condensed expression, at times obscure phraseology, and of the inaccuracies noted and the appearance, if not the fact, of confusion between general theory and specific application is to cover the trail for the ordinary statistical reader in his pursuit of the point of discussion and to compel him to stop and tread around in a disconcerting, and even discouraging, endeavor to find the trail again. He is little, if at all, interested in the making of trails on his own account in this field of exploration, or in cultivating ability do to so. He would like to make use of the road already opened by the Karl Pearsons, so far as he can learn it. But for that purpose, he wants to have the road shown to him and does not desire to reconstruct it.

By way of summary: This book is far more comprehensible to ordinary statisticians than any mathematical statement of statistical theory that the reviewer has seen. Anyone who has had algebra and who has the general notion of mathematical expression can follow most of the matter in the book if his interest in statistical theory is sufficient to withstand the resistance offered and if he has time enough outside of making his living to "plug" at the subject. What is here said in adverse criticism is not to be

taken as damning a book that is, in fact, at the head of the list for the purposes for which intended. It is merely to point out to non-mathematical statisticians that this is not *the* book for which they are waiting.

Indeed, the reviewer is inclined to think that that book cannot be written by a mathematician. His habits of thought and expression are against his success. Even if he could overcome that, his professional pride would stand in his way. That book must translate the mathematics of the subject into non-scientific language carrying over as few original expressions as possible. It will be a work, not of scholarship, but of adaptation of expression based on full knowledge of statistical practice. That would bring little credit to the author as a mathematician, and might well commit him to professionally childish work. That book will have to be written by one of the boneheads themselves who, by dint of perseverance, shall have learned the language of mathematics and then, having no mathematical reputation at stake, can dare to write with complete disregard of the highbrow standards of good form in mathematical discussion. His own full knowledge of statistical processes and his fellow feeling for his confrères will induce him to address them understandingly.

ARTHUR SARGENT FIELD.

Washington, D. C.

The Mathematical Theory of Population, of Its Character and Fluctuations, and of the Factors which Influence Them. Appendix A. Vol. I. Census of the Commonwealth of Australia. By G. H. Knibbs, C.M.G., F.S.S., F.R.A.S., etc., Commonwealth Statistician. Published under instruction from the Minister of State for Home and Territories, Melbourne, 1917. Pp. iv, 466.

The above monumental work represents a veritable milestone in statistical theory, which is emphasized by the fact that it has received almost universal praise from a number of leading statisticians the world over. Mr. Knibbs, the eminent statistician of the Australian Government, is by no means unknown to students of advanced statistical methods. His previous contributions have never been voluminous verbalisms, a very commendable but unfortunately rare occurrence in statistical literature. What they lack in quantity of paper they make up in originality and scholarship. The former small pamphlets from the hands of Mr. Knibbs have always brought to the present reviewer a very refreshing breeze of stimulating thought and a stream of new and original ideas. They have always been admirable in their brevity and lucid expression and stand in a class by themselves as examples of a clear and lucid exposition of thought.

One would therefore naturally expect something really valuable in this new tome of the Australian statistician. But even our fondest expectations were surpassed by the receipt of this massive volume of more than 460 large octavo pages, which for all times will stand as one of the leading works on mathematical statistics of the twentieth century. Not that the volume affords easy reading—on the contrary, I readily confess that I have not as

yet mastered several pages of the intricate analysis. But this shows the real character of the volume as something different from a mere text-book and marks it as a true piece of research work, which for many years to come will form a most valuable mine of information for future statisticians to work in. In that respect it is equal to the French classics on the calculus of probability by Laplace and Poisson, the works of Quetelet and the recent researches by Lexis and Bortkiewicz, Bachelier, Pearson, Edgeworth, Thiele, Westergaard, Charlier and Peano.

Knibbs, like his Scandinavian colleagues among the mathematical statisticians, shows a close familiarity with the leading works in various languages on the theory of statistics, the classical as well as the modern literature. While most English, American, French and German works on theoretical statistics have become hopelessly one-sided specimens of a particular school of statistical method, the unbiased and intimate knowledge of our Australian author adds greatly to the breadth and value of his work, at least for those who have a fairly good knowledge of international statistical literature. The footnotes show how busy a reader Mr. Knibbs has been of the foreign statistical literature. To students of small nations like those of the Scandinavian countries this becomes natural, because they must often publish their articles in two and often three journals of different countries of varying language to get a hearing before a larger body of readers than those of their own countries. The English writer on statistics does not encounter this obstacle (if it can be called so), and the knowledge of Mr. Knibbs of foreign writings in at least half a dozen languages is therefore very praiseworthy and probably accounts for his ability to select the most typical and important parts of such foreign writings as well as his own originality of handling his material. Take for instance the third chapter on "Curve Constants and Intermediate Values," and the following chapters on "Special Types of Curves," "Group Values" and "Summation and Integration," where the footnotes indicate that Mr. Knibbs has read nearly all the available mathematical literature and nevertheless been able to present a new and wholly original treatment, which especially becomes very instructive in his generalized flexible probability or frequency curves. Mr. Palin Elderton and Mr. Udny Yule, the well-known English statisticians, in their respective reviews of the present work in the *Journal of Institute of Actuaries* and the *Journal of the Royal Statistical Society* claim that the material has become bewildering in its detail and that our author has somewhat crowded out the methods of Professor Pearson and others in preference to his own. This I consider a picayune criticism and a very narrow-minded attitude. The very recent statistical methods as developed by my own Scandinavian compatriots, Charlier, Wicksell, Akeson and Jörgensen, and to which I also have contributed in a very modest measure, mainly following in the footsteps of Laplace and Poisson, are not mentioned by Knibbs. This fact should, however, not prevent one from giving unstinted praise to the excellent methods of Mr. Knibbs in deriving the equations of his flexible frequency curves.

In the chapter on nuptiality an equally interesting description of a

method of correlation is given in detail with the necessary mathematical formulaes and the protogamic surface.

The chapter on mortality is especially attractive. It is perhaps here more than anywhere else in the book that the author shows how advanced he really is in modern statistical matters. While other authors are content to harp on the old methods of constructing mortality tables along the well-known orthodox lines, Mr. Knibbs investigates the secular changes in mortality by a mathematical process and shows the necessity of constructing what he terms "fluent life tables." In the same connection he discusses also what he calls the "plasticity curve" as well as the composite character of aggregate mortality according to age and according to causes of death. Mr. Knibbs is here touching upon the same ideas which of late years have been taken up by the Scandinavian school of mathematical statistics, although by a different mode of treatment. He divides causes of death as regards *relative* frequency according to age in four typical groups.

In a brief space like this it is of course impossible to do full justice to a work planned on so large and broad a scale. A few selections from the foreword might, however, not be out of place. Mr. Knibbs says here that

"In the realm of official statistics there is an enormous amount of accumulated material, which decade after decade, remains unanalyzed and uninterpreted. This is due to several things, viz., to the fact that routine tabulations largely occupy the energies of the staffs of statistical bureaux, to the fact that much of the mass of material is defective and its correction involves more time than is available, and perhaps still more to the fact that appropriate schemes of mathematical analysis have as yet either not been developed, or are regarded as inapplicable."

In the present monograph Mr. Knibbs has materially advanced our knowledge of such mathematical methods and all earnest students will be thankful to him for it. The Australian Government also deserves high praise for having made possible the publication of this volume. We doubt whether public money has ever been expended in a better undertaking of statistical publications than the present work by the eminent scholar of the young and vigorous nation under the Southern Cross.

ARNE FISHER.

Harald Westergaard: "Statistikens Teori i Grundrids" (Outline of the Theory of Statistics). Second edition, completely rewritten, pp. 430, large octavo. Published through a grant of the University of Copenhagen by the C. E. Gad Publishing Company, Copenhagen, 1916.

This is one of the most important text-books on the theory of statistics that has appeared in recent years, and even a belated review of it may be welcomed by American readers for several reasons. Professor Westergaard is one of the leading statisticians of the world and has produced one of the finest treatises ever written on the theory of statistics. But the prime motive in penning this review is to correct certain erroneous opinions which we believe many English-speaking students are likely to form

in regard to the contents of this work. Curiously enough, such opinions about the exact character of Westergaard's book seem to have been caused chiefly by two recent articles from the pen of the same author: "Scope and Methods of Statistics" which appeared in the September number of this QUARTERLY, 1916, and a later one, "On the Future of Statistics," published in the May number, 1918, of the *Journal of the Royal Statistical Society.* These two very able articles leave one perhaps with the feeling that the distinguished Danish scholar is opposed to mathematical methods in statistical analysis, and for this reason they might be used (and in fact they have been used) as a very effective weapon against the small group of students who valiantly struggle—often against great odds—for the treatment of statistical mass data by mathematical formulae. Nothing, however, can be farther from the truth than the assumption that Professor Westergaard views mathematical statistics askance.

Westergaard is unquestionably one of the foremost leaders and advocates of the application of mathematical methods to the field of statistics; and his present work teems with mathematical formulae that require not only a knowledge of elementary algebra but also an extensive study of the differential and integral calculus. Readers who turn to "Statistikens Teori" in the belief that it is an exposition along the well-known lines of the German school of von Mayr and his American followers, like Mayo-Smith and H. Secrist, will therefore be disappointed.

Westergaard calls the theory of mathematical statistics an absolutely necessary "Tjener" (servant) in a number of investigations relating to so widely different subjects as social science, economics, medicine, heredity, biology, psychology, psychophysics, astronomy, physics and chemistry. The enormous development of statistics and statistical methods arises from the fact that the practical application of such methods lay hidden in the mathematical reasoning of a Bernoulli and a Laplace. It thus appears that mathematical statistics instead of being a mere foundling must be regarded as a science of old and renowned ancestry, dating back to a period when statistical data were very scant if at all available.

In a brief and lucid introduction the author characterizes the different statistical problems and then in the next three chapters deals with the history of statistics under these headings: "Statskundskab" (Science of Government), "The Political Arithmetic and Mathematical Theory of Probability," and "Modern Statistics." These three chapters contain a compact and at the same time complete description of the essential development of statistical methods down to the present time. In the fifth chapter the treatment of the collected data and their preliminary classification and analysis are described. This chapter deals with non-mathematical methods along the lines followed in the elementary works of Bowley and the newly issued text-book by Secrist. In the sixth chapter, which is entitled "The Regularity in Games of Chance and Similar Observations," Westergaard leaves the elementary and arithmetical methods behind and plunges into the application of the theory of probabilities.

Chapter VII continues with the development of the exponential curve

of the Laplacean-Gaussian Normal Curve of Errors derived from the point binomial by the methods of Laplace and Poisson. Here it becomes absolutely clear that Westergaard by no means ignores the mathematical formulae. Pages 130–162 teem with signs of integration and symbols from higher mathematics. The same exclusive mathematical treatment is continued in the eighth and ninth chapters under the headings, "Application of the Exponential Formula in Anthropometry" and "The Exponential Formula and the Manifestations of Life in Human Society." In these chapters Westergaard points out the necessity of testing whether the formulae from the calculus of probability can be applied directly to the statistical series. In other words, the question is of the stability of the statistical series tested by means of the mathematical theory of probability.

To the present reviewer it seems, however, that Professor Westergaard instead of using the somewhat lengthy tests by means of the normal error curve would have reached his goal quicker by aid of the so-called Lexian-Charlier dispersion theory. When one compares the striking illustrations of this theory in the recent work, "The Mathematical Theory of Probability and Its Application to Frequency Curves and Statistical Methods" (Vol. I, New York, 1916), by Professor Westergaard's colleague and compatriot, Mr. Arne Fisher, the superiority in some respects of the treatment of Mr. Fisher over that of his older and renowned countryman becomes evident. The Charlier coefficient of disturbancy, which is computed without much trouble, affords a very sensitive and powerful criterion for determining the possible presence of disturbing influences in the collected data. Charlier and Fisher have also given a simple mathematical method for eliminating secular perturbations in the collected data. It is to be regretted that these methods are not mentioned in Westergaard's book.

The eleventh chapter contains an excellent representation of interpolation and graduation methods. Those who have read the treatment of the same subject by Bowley will welcome the lucid and striking mathematical demonstration of the various formulae derived from the very general formula of Newton, at the hands of Westergaard.

The chapter on frequency curves shows that the Danish statistician keeps abreast of the modern mathematical development. Here he deals with Pareto's Law, the Curve of Utility, the Makeham Mortality Curves, and the frequency curves in their narrower sense. His treatment is too brief, however, to be of much practical value in actual calculations. Moreover, an error has crept into his demonstration of the computation of the parameters by means of Thiele's half invariants.

Finally, Westergaard discusses the theory of correlation based upon the method of moments and least squares in a most acceptable manner. This part fully exposes the error made by Professor West in the September number of this QUARTERLY, 1916, when he stated that Professor Westergaard apparently was not familiar with the properties of the correlation ratio. The whole trouble with the method of correlation lies at present in the fact that its mathematical theory is as yet unfinished. In this field the modern Scandinavian school of mathematical statisticians, such as C. V. L. Char-

lier, Sven Wicksell, N. R. Jörgensen and Arne Fisher is at present doing a veritable yeoman's service, and there is no doubt that before long we shall have a sound theory to build upon. The reviewer fully agrees with Westergaard that before this theory is finished much energy will be wasted through the efforts of some American economists in the application of the incomplete theory to economic problems. The mathematicians, however, are not to blame for blunders committed by economists who insist upon using the tools of an unfinished and incomplete mathematical theory to various complicated economic problems. A dull saw does not work very well, especially in knotty pieces of wood.

In others chapters Westergaard treats numerous practical problems from the fields of economics and insurance, such as the theory and use of index numbers, laws of wages, price theories and the theory of risk. Many of these chapters have become small gems of concentration and clear exposition of thought. The Danish statistician has here, on 6 to 7 pages, imparted more real knowledge than some of his American colleagues have spread over more than 80 pages of mere verbalisms.

The reviewer warmly recommends Professor Westergaard's splendid book on mathematical statistics. He is fully convinced, moreover, that the apparent criticisms by Westergaard of the advocates of ultramathematical methods in statistics has reference only to those who—to borrow a phrase from another eminent Danish scholar and statistician, the biologist Johannsen—habitually "treat statistics *as* mathematics and not *with* mathematics." The erroneous view which Westergaard's recent articles may have created in certain American circles arises solely from the fact that what he is apt to regard as elementary statistics are considered as higher mathematics by the majority of his American colleagues, and what he conceives of as higher mathematics is quite beyond the comprehension of most of our present day statisticians.

T. W.

COMMITTEE WORK IN FEDERAL STATISTICS AND LEGISLATION FOR THE FOURTEENTH AND SUBSEQUENT DECENNIAL CENSUSES.

In a paper read at the meeting of this Association at Princeton, N. J., December 29, 1914, Dr. S. N. D. North, Director of the Census from 1903 to 1909, suggested* that economists and statisticians take counsel of each other as to what course they could best pursue to promote the interests of governmental statistics, especially in view of the approaching Federal Decennial Census of 1920. Dr. North's paper was a continuation of part of the theme of Professor Willcox's article in the *Political Science Quarterly* for September, 1914. The Executive Board of our Association, acting upon this and other suggestions, appointed a Committee on Federal Statistics to coöperate with a similar committee of the American Economic As-

*"The Census Office in Commission." QUARTERLY PUBLICATIONS of the American Statistical Association, Vol. XIV, p. 467

sociation, and empowered that Committee to consider the whole problem of federal statistics in relation to the needs of the American economists and statisticians. The question of the 1920 Census was subordinate, therefore, to the problem of governmental statistics considered in its entirety.

Other associations and societies, actuated directly and indirectly by the suggestion of Dr. North, and having more specialized interests, appointed committees to consider the bearing of plans for the 1920 Census upon the subjects of inquiry with which they were concerned. These Associations and Societies were:

American Public Health Association, Section on Vital Statistics: Committee on Relation of 1920 Census to Vital Statistics, W. H. Guilfoy, Chairman; Edwin W. Kopf, Corresponding Secretary (July, 1916);

American Association for Labor Legislation; Committee on 1920 Census, Irving Fisher and John B. Andrews (February, 1917);

Casualty Actuarial and Statistical Society; Committee on Social Statistics and the 1920 Census, I. M. Rubinow, Chairman (April, 1917).

Representatives of the American Economic Association, American Statistical Association and American Public Health Association* met at the City Club, New York City, on January 27, 1917, discussed some of the problems of federal statistics, incidentally the 1920 Census, and adopted the following resolution:

> "That the coöperating committees favor the appointment by the President of a statistical commission to make a survey of federal work in statistics and to make recommendations for the coördination and improvement of such work."

The work of the American Public Health Association Committee has been outlined elsewhere.† The Committee appointed by the Casualty Actuarial and Statistical Society has been prevented from holding any meeting by war exigencies, but has had correspondence with the Director of the Census and with the Chairman of the Committee on Census, House of Representatives, chiefly with respect to Section 8 of the proposed bill. A round-table discussion of 1920 Census problems was conducted under the auspices of the American Statistical Association, December 28, 1917, at the meeting in Philadelphia. Mr. Koren presided. Dr. Joseph A. Hill, Dr. Helen L. Sumner, Dr. F. L. Hoffman, Dr. J. W. Trask and Mr. E. W. Kopf presented the views of federal bureaus and of some American learned societies. A quotation from Professor Allyn A. Young's Presidential address to our Association, December 27, 1917, will also be of interest as bearing on the attitude of American learned societies toward the Census and other federal statistical enterprises.

"A committee of this Association, coöperating with committees of other associations, has been studying the problem of the organization of statistical

* W. C. Mitchell, W. A. Hathaway, John Koren, W. F. Willcox, T. S. Adams, F. L. Hoffman, E. W. Kopf, N. I. Stone.

† New York Medical Journal, October 6, 1917, and QUARTERLY PUBLICATIONS of the American Statistical Association, September, 1916 and December, 1917.

work in Washington. I am informed that it has given favorable consideration to proposals looking toward the creation of a central statistical commission. It is hoped that its efforts will help to secure tangible progress toward that end. But the functions of such a commission ought to include more than merely introducing some measure of coördination into the work of the existing Washington offices. It should make recommendations to Congress for the inauguration of new statistical activities, utilizing either established offices or new ones according to the circumstances of the particular case."*

Meanwhile, the Federal Children's Bureau had presented to the American Association for Labor Legislation, meeting at Columbus, Ohio, December 29, 1916, a statement of its needs for statistics of the American family. That Association adopted a resolution, urging the tabulation of the facts collected upon the schedules of the 1910 Census.† The Children's Bureau could use, and in the opinion of responsible private statisticians should have, also, similar data from the records of the 1920 Census. Section 8 of the Bill as passed by the House, July 3, 1918, does not specifically provide for the enumeration of the facts. (See later comment on Section 8.)

The Director of the Census appointed a committee of experienced bureau officials on July 10, 1917, "to consider what changes in the act of July 2, 1909 . . . would be necessary or desirable in making provision for the next or Fourteenth Census." This committee held nineteen meetings between July 10 and December 4, 1917. The Thirteenth Census act was gone over, section by section, and only such changes were recommended as were, in the opinion of the committee, needed to insure efficiency and economy in the organization of the work of taking the Fourteenth Census. No radical revision of the Thirteenth Census act was deemed necessary. The importance of promptly completing the work of the Fourteenth Census within the limit of time allowed by law, or before July 1, 1922, was constantly kept in mind. To further these main ends, the Bureau Committee made due allowance for increase in the field and office forces, and for such increases in compensation as were believed to be necessary in order to retain or secure the grade of employees required for the work; it also aimed to avoid burdening the schedules with questions of little value, or of value not commensurate with the cost of securing the data. Certain defects in the Thirteenth Census act in connection with the prosecution of frauds were also remedied. The Director of the Census presented this report to the House Committee on the Census on February 20, 1918. The House Committee held numerous meetings between that date and April 10, 1918. The complete text of the bill submitted by Mr. Rogers, together with the discussion by the House Committee, was published, with index, in a document of 260 pages.‡

* Quarterly Publications of the American Statistical Association, March, 1918.

† American Labor Legislation Review, March, 1917, p. 187.

‡ "Hearing before the Committee on the Census, House of Representatives, Sixty-fifth Congress, Second Session, on a Bill to Provide for the Fourteenth and Subsequent Decennial Censuses." Government Printing Office, Washington, D. C., 1918.

The bill, amended in a few minor particulars as compared with the Census Bureau Committee draft, was introduced in the House on May 8, 1918, as H. R. 11,984 (*Cong. Rec.*, 65th Cong., 2d Session, p. 6712), and was referred to the Committee on the Census. The House Committee reported the Bill on May 21, 1918, without further amendment, accompanied by report No. 581. A minority report, No. 581 Part 2, was submitted by Mr. Hersey. On June 22 the House, sitting as a Committee of the Whole on the State of the Union, commenced work on the Bill. Mr. Hersey addressed the House on the general scope and objects of the proposed law. He urged that the 1920 Census be limited to an enumeration of the population (as required by Section II of Article I of the Constitution). Furthermore, Mr. Hersey objected in particular to the provision in the reported bill for "fifteen statistical experts at $2,000 each" and to the change in the method of appointing the supervisors of the Census. The bill reported by the House Committee stipulated that these supervisors be appointed "by the Secretary of Commerce upon the recommendation of the Director of the Census." Under former practice, these supervisors were appointed "by the President, by and with the advice and consent of the Senate." This objection brought in another question. The Bureau Committee had recommended, and the House Committee had so reported it, that the date of the Census should be advanced from April 15 to January 1, 1920. The change in the method of appointing the supervisors had been made in the draft submitted by the Bureau Committee, "because the Senate in all probability will not be in session at the time when the appointment of supervisors must be made. The next Census . . . will be taken in January, 1920. The present Congress will expire March 4, 1919 and the next Congress will normally not meet until December of that year. But the supervisors should be selected and commissioned between July 1, the date on which the Bureau is organized on the decennial Census basis and October 1, or ninety days before the date of the Census."* The change in the date of taking the Census (from April 15 to January 1) in Section 20 was made to meet the requirements of the Census of agriculture.†

Mr. Aswell replied to Mr. Hersey (p. 8813 of the *Record*), emphasizing the need (1) for taking the Census as of January 1, 1920; (2) for including a Census of agriculture, manufactures and of mines and quarries; (3) for revising the Thirteenth Census law so as to allow sufficient time for planning the 1920 enumeration, thus to avoid the initial haste and confusion which characterized the Thirteenth Census. After indicating clearly the character and provisions of each section of the Bill, Mr. Aswell considered some broad needs of the Bureau of the Census. His statement of the first need will be of interest to American statisticians who have had the larger problem of governmental statistics in mind.

* P. 21, Hearing before the Committee on the Census, February 20, 1918.
† Idem p. 28.

"Permanent recognition by Congress that the Bureau of the Census is the one statistical service of the National Government, created by the Congress for this sole purpose, to which should be assigned all inquiries for statistical facts required by this national body, as well as statistical work which is now being done in many different ways and more or less accurately by boards, commissions and a few other bureaus of departments of the regular service. We should remember that we created this service to do our statistical work and that we have made our annual appropriations to enable this office to thoroughly equip itself with statistically trained employees and all other resources and machinery for prompt, efficient and accurate statistical work. . . . Because of the decentralization of the statistical work of our National Government, no one of us today knows just where to secure the data we need. The evil has been growing for a number of years. The time has come for the National Congress to place itself on record in this matter. We should say to all branches of the national service in this city that the statistical work of our Government must be done by the statistical bureau we created to do this class of work."

The remainder of the debate of June 22 was devoted to further discussion of the advisability of changing the date of the Census, of incorporating a Census of agriculture, manufactures, mines and quarries. Debate on the sections of the bill began June 25, 1918. As finally passed by the House the various important sections are substantially as follows:

Section 1: The Census shall include: population, agriculture, manufactures, and mines and quarries in the United States, including each State, the District of Columbia, Alaska, Hawaii and Porto Rico. A Census of Guam and Samoa shall be taken in the same year by the respective governors of those islands and of the Panama Canal Zone by the governor of the Canal Zone in accordance with plans prescribed and approved by the Director of the Census.

Section 3: In addition to the staff provided by the Legislative, Executive and Judicial appropriation act for the fiscal year immediately preceding the decennial Census period, there may be employed: an Assistant Director, a chief statistician, a disbursing clerk, an appointment clerk, a private secretary to the Director, 4 stenographers, 8 expert chiefs of divisions, and 15 statistical experts. The Assistant Director shall be appointed by the President, by and with the advice and consent of the Senate. The chief statistician, the disbursing clerk, the appointment clerk, the chiefs of divisions, and the private secretary to the Director shall be appointed without examination by the Secretary of Commerce upon the recommendation of the Director of the Census. The statistical experts and the stenographers shall be appointed in conformity with the civil service act and rules.

Section 5: Provides salaries of officials as follows:

Director	$7,000
Assistant Director	4,500
Three chief statisticians and the chief clerk	3,000 each

Two other chief statisticians	3,000 each
Geographer	2,750
Disbursing clerk	3,000
Appointment clerk	2,750
Chiefs of divisions	2,250 each
Private secretary to the Director	2,250
Statistical experts	2,000 each
Stenographers provided for in Section 3	1,800 each

Section 8: The heart of the bill. Stipulates the topics or inquiries to be included in the Census schedules. The population schedule is to include queries for each inhabitant as to: name, place of abode, relationship to head of family, color, sex, age, conjugal condition, place of birth, place of birth of parents, nationality or mother tongue of all persons born in foreign countries and of parents of foreign birth, number of years in United States, citizenship, occupation, whether or not employer or employee, whether or not engaged in agriculture, school attendance, literacy, tenure of home, and the name and address of each blind or deaf and dumb person. The agriculture schedule is to include name, color, sex, and country of birth of occupant of each farm, tenure, acreage of farm, acreage of woodland, value of farm and improvements, the encumbrance thereon, value of farm implements, number of live stock on farms, ranges and elsewhere, and the acreage of crops and the quantities of crops and other farm products for the year ending December 31 next preceding the enumeration. Inquiries to be made also as to the quantity of land reclaimed by irrigation and drainage and the crops produced; also the location and character of irrigation and drainage enterprises and the capital invested in such enterprises. Manufactures, and mines and quarries schedules shall include name and location of each establishment, character of organization, whether individual, corporate or other form, character of business or kind of goods manufactured, amount of capital actually invested, number of proprietors, firm members, copartners and officers, and the amount of their salaries, number of employees and the amount of their wages, quantity and cost of materials used in manufactures, principal miscellaneous expenses, quantity and value of products, time in operation during year, character and quantity of power used, character and number of machines used.

The American Public Health Association Committee addressed a letter to the Director of the Census with respect to the inquiries on the population schedule. The letter asked for an expression of Bureau opinion on the inclusion of the following inquiries on the population schedule, the Committee having borne in mind the need for restricting the scope of the Census:

1. The American Public Health Association Committee had received several requests from its correspondents that it incorporate in its communication to the Director on the legislation for the 1920 Census, the subject of family statistics. Specifically the Committee's letter asked that for each married or widowed person the following facts be enumerated:

(a) Calendar year of marriage (or, number of years married).
(b) Number of children live-born to the marriage.
(c) Number of such live-born children living on Census Day.

The Director replied substantially as follows:

Inquiry regarding the number of children born and the number living on Census Day had been made part of the Eleventh Census (1890) and of the Twelfth Census (1900). No tabulations were made of the 1890 data, and but limited use was made of the 1900 facts (28th volume of the Immigration Commission report). The Thirteenth Census schedules bore the following inquiries in regard to each married, widowed or divorced woman, "number of years of present marriage" and "mother of how many children," the latter question being sub-divided into (1) number born, *i. e.*, the number of children (excluding stillborn) that each such woman had had during her lifetime, by her present or by any former marriage, and (2) the "number of children now living," whether they were residing with her or elsewhere. Various practical difficulties prevented the issue of any tabulations of these subjects of inquiry. Such a tabulation had been tentatively provided for as a part of the Bureau program, but this is now contingent somewhat upon the demands upon the Bureau for war work. "The inclusion in the schedule for 1920 of inquiries similar to those for 1910 also can be decided later, as the form and subdivision of the inquiries to be made on the various Census schedules is determined by the Director of the Census, and the matter will thus receive due consideration at the proper time."*

2. Whether person is a survivor of the war with the Central Teutonic Powers. The Director was asked whether it would be advisable to provide for an enumeration of survivors of the present war. The Director's letter in reply suggested that consideration of the possibility of providing for this enumeration be deferred until the period for taking the Census is nearer at hand and the probable status of affairs on Census Day can be more nearly determined.

3. Interdecennial or State Censuses. The Committee suggested that a provision for Federal subsidy of State or Interdecennial Censuses, similar to Section 22 of the Census Act of March 3, 1879, be included in the 1920 Census Bill. The Director's letter said that the Bureau was entirely mindful of the need for a more frequent enumeration of population. It was not thought practicable to attempt to secure authority for an interdecennial Census of population of limited scope even, as a part of the Act providing for the Fourteenth Decennial Census. In the opinion of the Bureau the necessary legislation should not be sought until after the substantial completion of the decennial enumeration of population and the publication of the reports early in 1922 is assured.

Section 9. Provides for appointment of 26 supervisors. See comment in preceding part of this note on objections on the floor of the House to mode of appointing the supervisors. (Pages 9136–9143, *Cong. Record*, 65th Congress 2d Session, give full text of debate on this section.)

* See page 7, "Tentative Program of the Bureau of the Census, 1916–1919," Government Printing Office, Washington, D. C., July, 1916.

Section 12. Stipulates how and when the Census shall be taken. Census date for enumeration of population changed from April 15 to January 1. See also Section 20.

Section 23. A new paragraph has been inserted, the purpose of which is to eliminate, as far as possible, the "boosting" tactics employed in many cities and towns during the taking of the last decennial Census.

The bill passed the House July 3, 1918. The matter now rests with the Senate Committee on the Census. A further note will be submitted on the course of the bill in the Senate(

E. W. KOPF.

American Public Health Association Committee on Relation of 1920 Census to Vital Statistics, New York City.

Skandinavisk Aktuarietidsskrift. 1918, No. 1–2. Stockholm.

About the year 1903 there appeared a remarkable little journal, entitled "Aktuaren" (The Actuary), published by Scandinavian actuaries. The number contained a series of brilliant articles, showing how advanced the Scandinavian actuaries were in their profession. However, the promising beginning resulted only in a single number. The political differences between Norway and Sweden which terminated the personal union between the two countries killed the attempt to produce a common organ for the actuaries of three sister nations.

In 1914 the Swedish Actuarial Union launched a new journal, entitled "Svenska Akturarieföreningen Tidsskrift." It is this journal which now has become the common organ under the above name, for all the actuaries and statisticians of the three countries. It would, however, be erroneous to regard it as an exclusive Scandinavian journal. In fact, it is truly cosmopolitan and international in character and scope. The present number contains articles in French, English, Danish and in German, the latter article in German being contributed by the eminent Russian statistician, Bortkiewicz. Other articles in English by the Danish actuary, Dr. Steffensen, and the eminent young Swedish statistician, Dr. Wicksell, show how rapidly the Scandinavians are out-distancing their colleagues in other countries in researches on modern mathematical statistics. There is no doubt that the labors of the ultra mathematical statistical school of Scandinavia, under the able leadership of such eminent scholars as Charlier, Wicksell and Jorgensen, have surpassed even the English biometricians in elegance and general application of their methods. It is, however, doubtful if many of our American statisticians and economists would benefit much by the Scandinavian researches, because such researches require an extensive training in higher mathematical analysis, which, unfortunately, our actuaries and statisticians as a rule do not possess. It is a significant fact that the University of Lund has established a professorship in Mathematical Statistics. While we have had plenty of chairs of statistics in our universities this is probably the first instance of a department that devotes its time exclusively to mathematical statistics. The present occupant of the

chair is Dr. Wicksell and we doubt whether the Swedish government could have made a better choice than that in giving the professorship to this brilliant young scholar whose recent researches on the theory of probability rank among the foremost of the present time.

A. F.

Is War Diminishing? By F. A. Woods and A. Baltzley. Boston: Houghton Mifflin Co., 1915.

The present deluge of war books, relating directly to the war or social questions growing out of the war makes it rather difficult to criticize or review incisively any one specific work. It will be interesting to see how many of the present war books will survive when once the minds of men settle down to normal conditions. Evidently the vast majority of the present books on the war will soon be forgotten and only a few will remain. To the present reviewer one of the most interesting war books is a little brochure written by a Chinese scholar. This deep thinker of the Orient points out the futile attempt of many writers to place the blame for the war on certain political parties. The Occidental mind is of a too prospective turn, always looking to the future, and neglects often to compare present conditions with the past. As the Chinese scholar points out, it is of greater importance to compare our own conduct with that of our forefathers than to speculate on what the coming generations will think of us. The white race certainly has no reason to feel superior or proud in comparison with the yellow race when it has reared such a monster as the German Empire. To the Chinese scholar the whole question of the war is that of sex, this term being taken in its most general sense. He divides the peoples into "he" and "she" nations. Germany is an essentially feminine nation according to the opinion of this wise Oriental, who evidently has looked deeper below the surface of the different nations of Europe than any other writer of modern times. England and France, on the other hand, he regards as nations with typically masculine traits.

After reading the very stimulating book by the Chinese scholar, the reviewer began to see in a new light some of the interesting ideas in the book by Dr. Woods. Dr. Woods has not tried to write a brief for any one of the various participants in the present war. He has followed a purely retrospective course and limited his research to the concrete and narrow question: "Is War Diminishing?" This severe limitation of his theme does not allow any room for the vague generalities indulged in by various pacifists and peace advocates. His very pertinent remarks on the utterances of such apparently great authorities as Ex-President Taft, Dr. Butler of Columbia University, Professor Thorndyke and Dr. Mitchell will warm the heart of every true scientist. Unfortunately vague generalities and mere verbalism have too often been indulged in by many of our leading citizens. In the past attractive presentation and cheap oratorical phrases have been preferred by the public to the less ornamental language of the scientist. The white man, and especially the German, has in the past

decade indulged in a lot of bombastic generalities, which after all have turned out to be mere bubbles of speech. While our "leading" men of the past have talked volumes about the horrors of war and held up before their audiences the chimera of eternal peace, they have never given deep thought to the experience of the past.

Dr. Woods suggests the measurement of the warlike tendencies of the different nations by the duration of the wars in which they have participated. In this respect he finds that Great Britain must be classed as an essentially warlike nation, because of its almost uninterrupted warfare. However, it must be remembered that the British Empire is founded upon maritime power and most of its wars have been fought chiefly on the sea. Naval actions do not require the large number of armed men involved in military operations on land; hence the drain on England will in all probability be found to have been much smaller than on the other nations. It is, however, almost impossible for any writer to measure the number of men who have been engaged in the past warfare of the different nations. For this reason the actual duration of the different wars must at the present serve only as a rough index of the secular variation of warfare.

In his introduction Dr. Woods touches upon many interesting questions, among other things the much debated subject of eugenics. But he does not adhere to the extravagant views of the ordinary "social uplifter" and disgruntled feminist who apparently sees a monster in every male. He is even of the opinion that war may have a beneficial influence upon a nation if it results in a selection of the best fitted survivors of an originally strong and healthy lot. His remarks that man has not evolved essentially by a raising of the average, but that the rise in intelligence of a very small percentage of all mankind has been the feature in the growth of civilization, are well worth notice. One of the most baneful influences of our modern intellectual life is its mediocrity. The average intellect undoubtedly has been advanced tremendously in most countries, but the age of intellectual geniuses of the type of Leonardo de Vinci, Descartes, Newton and Leibnitz seems to be a thing of the past. A few years ago poor benighted Russia seemed the only place in the white man's world from which occasional flashes of genius broke forth. Dr. Woods seemingly is of the opinion that man has evolved in spite of the fact that he has fought presúmably half of the time. If warfare therefore is so deleterious it may be asked, How did man get where he is? Of course, it may be a debatable question if man really has evolved. The reviewer is of the opinion that our present age has retrogressed in intellectual attainments in comparison with Greek antiquity and the Italian renaissance. The white man of the future will, in his opinion, have to relinquish the leadership to the yellow man, who in many respects possesses greater staying power and national stability than the white nations of Europe and America. Eventually this may prove of great benefit to mankind, and it seems to him sheer nonsense to speak of a yellow peril. The Chinese mind is in most respects much more dignified than our European mind, and the philosophy of the East is probably sounder than the often highly speculative systems of the Western world.

At any rate, any possible conquest by the Chinese or Japanese would be preferable to the "Weltmacht" of the German.

While the introductory notes by Dr. Woods are very suggestive and original, the present reviewer cannot bestow the same praise on the part of the work accredited to Mr. Baltzley. The statistical summaries and diagrams prepared from various treatises and text books on history are comparatively crude in their makeup. Nowhere does the reader find even the slightest attempt at a mathematical and quantitative analysis of the data. The chronological tables as prepared by Mr. Baltzley do not represent original research beyond that of a simple tabulation. Probably the author has something better in store for us, but if this be the case, he ought to have indicated it in the text. Mr. Baltzley's comments on Poland are of interest; but the downfall of the Polish power must be sought for in a quite different direction from that of pure military extinction. The writer believes it to be a fact that all of the former great nations of Europe, such as Spain, Portugal, Poland and Turkey began to decline as soon as the Jewish element exceeded a certain percentage of the total population; and history seems to be repeating itself in the case of the rapidly disintegrating Austrian Empire, in which the Jews, by their financial strength and by their proportional strength in population, have exerted a tremendous influence. This is a very interesting phenomenon which Dr. Woods and his collaborator might study with profit.

ARNE FISHER.

REPLY TO DR. DAY'S CRITICISM OF MY ARTICLE ON CORRELATION COEFFICIENTS.

It would be unwise to occupy too much of the valuable space of "The Publications" with a lengthy discussion, but Dr. Day's well-stated criticisms are so broad in scope that they seem to call for a brief reply.

He first dissents from my observation as to the frequency of the misuse of the coefficient by reputable statisticians. I hope that he is right and that the cases of that type coming under my observation are fewer than my memory indicates. The instances in question are not at hand at present, but vivid recollections come to mind in which economists who habitually present statistical studies to the public have used coefficients to prove correlation between the cyclical fluctuations of two variables without getting rid of the trend or the lag. However, the question as to the standing of writers misusing the coefficient does not seem to me to be one of importance. My criticisms were intended entirely for the benefit of those who did not use the coefficient correctly. Experts in correlation would, of course, not need any direction concerning its application or interpretation.

Statistical literature has, however, for example, been too commonly favored with contributions in which a coefficient has been used to bolster up an assumption of correlation between short-time fluctuations when the only reason for the high coefficient was the fact that both variables had

similar trends which had not been eliminated. In such instances, of course, the data would give a high positive coefficient whether the correlation between the shorter oscillations was direct, inverse, or entirely non-existent.

The tendency to assume that the high coefficient shown by two variables with similar steady trends is a proof of correlation between the two is probably the most flagrant misuse of the coefficient. As a matter of fact, Pearson's coefficient is entirely without significance for comparing historical variables unless there are a *considerable number* of corresponding up-and-down swings in the variables. If such corresponding swings occur, the plotted curves usually reveal the fact quite clearly. My own rule is that when the correlation is not entirely evident from the curves, it is unsafe to assume that correlation exists. When an examination of the curves leaves one in doubt, the coefficient, if computed correctly, will normally be very low. If the curves seem distinctly to indicate correlation, it certainly does no harm to clinch the evidence by appealing to a high coefficient to prove how successful one has been in eliminating the extraneous factors. It should, however, always be kept in mind that two trends are likely to be similar because of wholly unrelated causes; that in such instances, Pearson's coefficient will show very high correlation with a low probable error; and that, despite this apparently conclusive mathematical evidence, one is wholly unjustified in assuming any trace of interdependence or inter-relationship between the two variables. Two complex curves may be so similar that, with no other evidence whatever, one may be fully justified in concluding that they prove causal relationship. On the other hand, a large coefficient with a small probable error only establishes a similarity between the dominant tendencies of the two variables and, without exception, requires an examination of the plotted graphs or additional evidence from other sources before any assumption of a causal relationship is justifiable.

Dr. Day next takes exception to my definition of correlation. I still fail to see that I have in any way essentially changed the idea as it has been generally understood by statisticians. I have only sought to make definite and clear-cut a concept which is frequently described in a more or less nebulous and hazy manner. And it still seems to me that "a tendency toward persistent association," as he puts it, between varying phenomena, must be either chance or causal. The former may interest the mathematician—the latter only is of significance to the economist. In my opinion, the latter relationship is the only one in the mind of the economist who refers to correlation. If a given cause always produces a definite effect which proposition Dr. Day does not dispute, then correlation is always perfect no matter how hard it may be to isolate and detect it. The different blind men who described the elephant so diversely did not by their description alter by one iota the characteristics of the elephant.

I cannot agree that the fact that correlation is always perfect, if existent, makes the problem of demonstrating the relationship between two variables necessarily a simple one. Extraneous factors sometimes are hard to

eliminate, but only when these are disposed of successfully can the coefficient or any other method really prove correlation to exist. With other factors still present, a high coefficient indicates merely spurious correlation, due probably to the accident of similar trends.

The use of coefficients, therefore, is not in any sense a legitimate substitute for the elimination of the effects of factors other than those which it is desired to compare. If the troublesome interfering causes cannot be scientifically disposed of, to the required degree, then the use of coefficients gives no aid whatever, and we are compelled to concede defeat and pass the problem on to our successors.

I agree fully to "the practical impossibility of isolating economic forces *completely*." Neither are chemical or physical forces ever *completely* isolated. Nevertheless, it must be admitted that physical and chemical laws have been derived and established which are accurate enough to serve the ends in view. Identically similar conditions rule in the economic field except that the difficulties of isolation may, in some instances, be even greater than those besetting the physicist or chemist. I would be the last to assert the present possibility of solving correctly nearly all questions of correlation of economic facts, but fortunately my own experience with scores of typical economic problems demonstrates conclusively that it very often is perfectly feasible to eliminate all undesired influences to the extent necessary to show conclusively whether correlation does or does not exist. The process of segregation having been carried thus far, the facts are so evident that it is wholly immaterial whether they are presented in the form of coefficients or curves. To my mind, the most important single criterion of success for the statistical economist is his ability to isolate to the necessary degree the forces in which he is interested. The method of measuring these relationships is decidedly of secondary importance.

In conclusion, I wish to thank Dr. Day for his courteous comments. I also wish to give assurance that while my criticism concerning the use of the coefficient was not entirely confined to amateurs, I have had no thought of indicting all statisticians indiscriminately. I also concede freely the fact that the correlation coefficient, when intelligently used, is a perfectly satisfactory tool for measuring one's success in segregating a cause and its effect from a mass of confusing data.

WILLFORD I. KING.